4-10-2009

A Generalization of AUC to an Ordered Multi-Class Diagnosis and Application to Longitudinal Data Analysis on Intellectual Outcome in Pediatric Brain-Tumor Patients

Yi Li

A GENERALIZATION OF AUC TO AN ORDERED MULTI-CLASS DIAGNOSIS AND APPLICATION TO LONGITUDINAL DATA ANALYSIS ON INTELLECTUAL OUTCOME IN PEDIATRIC BRAIN-TUMOR PATIENTS

by

YI LI

Under the Direction of Yu-Sheng Hsu

ABSTRACT

Receiver operating characteristic (ROC) curves have been widely used in evaluation of the goodness of the diagnostic method in many study fields, such as disease diagnosis in medicine. The area under the ROC curve (AUC) naturally became one of the most used variables in gauging the goodness of the diagnosis (Mossman, Somoza 1991).

Since medical diagnosis often is not dichotomous, the ROC curve and AUC need to be generalized to a multi-dimensional case. The generalization of AUC to multi-class case has been studied by many researchers in the past decade. Most recently, Nakas & Yiannoutsos (2004) considered the ordered d classes ROC analysis by only considering the sensitivities of each class.

Hence, their dimension is only d. Cha (2005) considered more types of mis-classification in the ordered multiple-class case, but reduced the dimension of Ferri, at.el. from d(d-1) to 2(d-1).

In this dissertation we are trying to adjust and calculate the VUS for an ordered multiple-class with Cha's 2(d-1)-dimension method. Our methodology of finding the VUS is introduced. We present the method of adjusting and calculating VUS and their statistical inferences for the 2(d-1)-dimension. Some simulation results are included and a real example will be presented.

Intellectual outcomes in pediatric brain-tumor patients were investigated in a prospective longitudinal study. The Standard-Binet Intelligence Scale-Fourth Edition (SB-IV) Standard Age Score (SAS) and Composite intelligence quotient (IQ) score are examined as cognitive outcomes in pediatric brain-tumor patients. Treatment factors, patient factors and time since diagnosis are taken into account as the risk factors. Hierarchical linear/quadratic models and Gompertz based hierarchical nonlinear growth models were applied to build linear and nonlinear longitudinal curves. We use PRESS and Volume Under the Surface (VUS) as the criterions to compare these two methods. Some model interpretations are presented in this dissertation.

INDEX WORDS: ROC curve, Volume under the surface (VUS), Longitudinal data, Standard Binet Intelligence, Hierarchical linear model, Gompertz growth model, PRESS, Pediatric brain tumors

A GENERALIZATION OF AUC TO AN ORDERED MULTI-CLASS DIAGNOSIS AND APPLICATION TO LONGITUDINAL DATA ANALYSIS ON INTELLECTUAL OUTCOME IN PEDIATRIC BRAIN-TUMOR PATIENTS

by

YI LI

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2009

A GENERALIZATION OF AUC TO AN ORDERED MULTI-CLASS DIAGNOSIS AND APPLICATION TO LONGITUDINAL DATA ANALYSIS ON INTELLECTUAL OUTCOME IN PEDIATRIC BRAIN-TUMOR PATIENTS

by

YI LI

Committee Chair: Yu-Sheng Hsu

Committee: Tricia King
Gengsheng Qin
Xu Zhang
Yichuan Zhao

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
May 2009

ACKNOWLEDGEMENTS

I am highly indebted to my chair, Dr. Yu-Sheng Hsu, who manages to give me precious advice at critical times despite the very high demand for his time. I appreciate his allocation of time to participate in my research where he gave me his insightful suggestions of great consequence. I am also grateful for the opportunities hat he gave me as his GRA during my study. I'd like to express my appreciation for classes I took from him. I enjoyed every moment in the classes. My reason to go to Georgia State to pursue my Ph.D. was to study and do research with Dr. Hsu and I am pleased that I made that decision. I would like also to express my gratitude for Dr. Hsu and his family's thoughtfulness of including me in their Thanksgiving dinners. This has made me feel warm and less homesick.

I thank Dr. Tricia King sincerely for her great help and suggestion throughout the whole project. I thank my other committee members, Dr. Gengsheng Qin, Dr. Xu Zhang and Dr. Yichuan Zhao to spend time on reading this dissertation and provide precious comments.

I would like to thank Brains & Behavior Program and acknowledge the funding for my graduate study and research in Georgia State University.

I am grateful to my parents, for always supporting me in whatever I wanted to do. Without their love and support, I am sure I wouldn't imagine going this far. I would like to dedicate this final product to them.

I would like to thank the manuscript reviewer Tippi Hyde, who help me to review and correct my dissertation time and time again. Thanks for her time and patience.

I also thank all my friends in Atlanta for making my life here much more enjoyable. You have added a splash of color to what could have been a drab experience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction of ordered ROC analysis

In order to support a decision from researchers in the field of medicine, classification and its accuracy with a diagnostic test have been considered as one of the most important processes in the study. When their decisions or predictions are incorrect in a diagnosis of a patient, costs of misclassification could be enormous. Thus, both avoiding misclassification and increasing accuracy can be the goal in medical decision making.

Receiver Operating Characteristic (ROC) curve has been used as a tool in distinguishing the quality of given classifier in diagnostic tests for years. ROC analysis was developed to describe and summarize the data from electronic signal detection (Tanner & Swets, 1954; Egan, 1975). Nowadays, it has been used in the biomedical applications to evaluate the accuracy of diagnostic and prognostic technology (Hand & Till, 2001, Mossman, 1991).

The area under the ROC curve that can be calculated with sensitivities and specificities at various cutoff points is used for the evaluation of the misclassification. Sensitivity is defined as the proportion of cases correctly identified by the test among all positives (e.g. the proportion of positive tests among all disease patients), and specificity is defined as the proportion of cases correctly identified by the test among all negatives. ROC curve is a graph of sensitivity versus one minus specificity at all cutoff points. According to James Hanley (1982), the area under the ROC curve (AUC) implies a single quantitative index to reduction of entire ROC curve to quantify a diagnostic accuracy with the information of entire ROC curve. However, most of ROC analysis and AUC have been restricted to a classifier with just two classes.

However, we are often encountered with the medical problems in which there are more than two classes. Srinivasan (1999) extended three classes problem to 6-dimensional ROC surface using the six misclassification cells as the six dimensions. To compute the volume under the ROC surface (VUS), we need to find the volume of the convex hull generated by n points in the six dimensional spaces. Using some existing computer software, we can compute the volume of the convex hull. However, it is very easy to miss some points in constructing the convex hull due to these high dimensional surfaces, i.e. six dimensions in three classes. Hand & Till (2001) used the average of all possible paired comparisons to extend the case of more than two classes and reduce their measure to the standard form in the two classes. In addition, Mossman (1999) presented an approximation for the three classes based on a priorisation of one of them. Ferri, Hernandez-Orallo, and Salido (2003) [1] proposed a complete solution of multi-dimensional problems of ROC analysis using again all misclassification cells as dimensions. They presented the trivial classifiers, the maximum and minimum of the VUS in the extension of AUC to the three and more classes. They introduced the method of Hyperpolyhedron Search Algorithm (HSA). However, their paper, also, has shown the theoretical limitation of computing the maximum volumes in the high-dimensional problem and at least d(d-1) dimensional variable for d classes are needed for obtaining their volumes. All the above works are treating each class equally. Although the costs for each misclassification can be assigned, they never can be embedded in the structure of VUS. In other words, they can not be applicable for d ordered classes.

Most recently, Nakas & Yiannoutsos (2004)[2] considered the ordered multiple-class ROC analysis by only considering the sensitivities of each class. Hence, their dimension for obtaining their volumes is only d for d classes. However they did not consider the specificities.

Cha (2005) [3] considered more types of mis-classification in the ordered multiple-class case, but reduced the dimension of Ferri, at.el. to 2(d-1). Also she uses Qhull method to compute the multi-dimensional volume. The computation can be time consuming and has no mathematical expression for the estimation. Also, no inferences were derived.

In the first five chapters of dissertation, we are trying to define and calculate the VUS for an ordered multiple-class with Cha's 2(d-1)-dimension method. In chapter 2, our methodology of definition and estimation of the VUS is introduced in 3 classes' case. Also a short example will be stated in this chapter. We present the method of making inference to VUS for 3 classifier case in chapter 3. We extend our conclusion to any number of classifiers in chapter 4. Some simulation results are included in chapter 5. The chapter 6, 7 and 8 will present a longitudinal data analysis with application of VUS. The last chapter will give the conclusion and future research.

CHAPTER 2

New VUS definition and estimation

2.1 Previous research conclusion

We start with a 2-class diagnosis case.

Sensitivity and specificity are used to evaluate the accuracy of a diagnostic or classificatory tool. Sensitivity is the proportion of correctly diagnosed positives among all true positives. Specificity is the proportion of correctly diagnosed negatives among all true negatives (Walld, 2001). These values can be illustrated in the following tables:

Table 2.1 2-class diagnosis table

		Diseased Status	
		Positive	Negative
Test	Positive	#n TP (True Positive)	#n FP (False Positive)
	Negative	#n FN (False Negative)	#n TN (True Negative)

Then we have

Sensitivity=P (diagnosed Positive| Disease)

Specificity=P (diagnosed Negative| Non-disease)

Let X_i be the continuous random variable used to diagnose a certain disease with $i = 1$ representing it was from the disease class and $i = 2$ representing the benign class, respectively. The sensitivity and the specificity can be expressed as

$Sensitivity = P(X_1 \le C)$ and

$Specificity = P(X_2 \ge C)$ for a cutoff point C.

Bamber (1975) [4] proved that

$$AUC = \int_0^1 sensitivity \; d(1 - specificity)$$

$$= \int_0^1 P(X_1 \le C) dP(X_2 \le C)$$

$$= P(X_1 \le X_2) \qquad (2.1.1)$$

We can also derive the AUC by expectation. First, let X_i be the diagnosis result (fix number) of a certain disease with $i = 1$. The sensitivity and the specificity can be expressed as

$Sensitivity = I(X_1 \le C) = U_1$ and

$Specificity = I(X_2 \ge C) = U_2$ for a cutoff point C,

where I is the indicator function.

Similar to Bamber's proof,

$$h(X_1, X_2) = \int_0^1 \int_0^{U_1(C)} dU_1 dU_2 = \int_0^1 U_1 dU_2$$

$$= \int_{C \in (-\infty, +\infty)} I(X_1 \le C) dI(X_2 \ge C)$$

$$= I(X_1 \le X_2) \qquad (2.1.2)$$

Then we take expectation to $h(X_1, X_2)$ we can get

$$E_{(X_1,X_2)}h(X_1,X_2) = P(X_1 \le X_2) = AUC \tag{2.1.3}$$

We call $I(X_1 \le C)$ and $I(X_2 \ge C)$ the kernel of sensitivity and specificity separately and $h(X_1,X_2)$ the kernel of AUC.

Let $\{X_{1i} \mid i = 1,....n_1\}$ and $\{X_{2j} \mid j = 1,....n_2\}$ be two random samples from "disease" and "benign" populations, respectively. Then the estimated AUC can be expressed as

$$\hat{AUC} = \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} I(X_{1i} \le X_{2j}) \tag{2.1.4}$$

This scheme can be generalized to 3-class case as follows: Let the three classes be "Disease (D)", "Suspicious (S)" and "Benign (B)". Also, let X_i be the classifier in three populations with i=1, 2 and 3 representing D, S and B respectively. If we use C_1 and C_2 as the cutoff points ($C_1 \le C_2$), then we have similar table as 2-class case, where we use "h" as sensitivities and "F" as mis-classified probabilities.

Table 2.2 3-class diagnosis table

		Actual diseased Status		
		Disease	Suspicious	Non-Disease
Test	Positive	h_a	F_1	F_2
	suspicious	F_3	h_b	F_4
	Negative	F_5	F_6	h_c

Then we have

$$\text{Sensitivity}_1 = \text{P (diagnosed D | D)} = P(X_1 \le C_1) = E(U_1), \tag{2.1.5}$$

$$\text{Sensitivity}_2 = \text{P (diagnosed S | S)} = P(C_1 \le X_2 \le C_2) = E(U_2), \tag{2.1.6}$$

$$\text{Sensitivity}_3 = \text{P (diagnosed B | B)} = P(X_3 \ge C_2) = E(U_3), \tag{2.1.7}$$

Here, $U_1 = I(X_1 \le C_1)$, $U_2 = I(C_1 \le X_2 \le C_2)$, and $U_3 = I(X_3 \ge C_2)$ are the kernels.

If we ignore the various mis-classifications, i.e. no specificities are concerned, then the VUS can be expressed as, see Nakas & Yiannoutsos (2004):

$$VUS = \iint_{C_1 \le C_2} Sensitivity_2 \, dSensitivity_1 \, dSensitivity_3 = P(X_1 \le X_2 \le X_3) \tag{2.1.8}$$

Also, we can first considered X_1, X_2, X_3 as fixed diagnosis for the three classes and use the kernels of the sensitivities to derive the kernels of VUS and take expectation to find VUS

$$h(X_1, X_2, X_3) = \iint_{C_1 \le C_2} U_2 \, dU_1 \, dU_3 = I(X_1 \le X_2 \le X_3)$$

$$VUS = E_{(X_1, X_2, X_3)} h(X_1, X_2, X_3) = P(X_1 \le X_2 \le X_3) \tag{2.1.9}$$

The estimate of VUS can be expressed as

$$\hat{\delta} = U_{n_1, n_2, n_3} = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} h(X_{1i}; X_{2j}; X_{3k}) = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} I(X_{1i} \le X_{2j} \le X_{3k}) \tag{2.1.10}$$

where $\{X_{ij} \mid j = 1, ..., n_i\}$ and i=1, 2, 3 are three random samples from D, S, and B population, respectively.

The U-statistics kernel is $h_1(X_1; X_2; X_3)$.

Using the fact that $E(h_1(X_1, X_2, X_3)) = VUS$ and applying the Central Limit Theorem, we have

$$\sqrt{n_1 + n_2 + n_3}\,(U_{n_1,n_2,n_3} - VUS) \xrightarrow{L} N(0, \sigma^2)\,. \tag{2.1.11}$$

2.2 New proposed method

If all six different mis-classifications are considered, then the surface is in a six dimensional space. For an ordered multiple-class case, we propose a four dimensional surface as follows:

Sensitivity_1=P (diagnosed D | D) (2.2.1)

Specificity_1=P (diagnosed S or B | S or B) (2.2.2)

Sensitivity_2=P (diagnosed D or S | D or S) (2.2.3)

Specificity_2=P (diagnosed B | B) (2.2.4)

This setup is combining P(diagnosed D| S) and P(diagnosed D| B) into P(diagnosed D| S or B), and the method was motivated by assuming the following usual facts:

P(diagnosed D| S or B)>>P(diagnosed D| B), P(diagnosed B| S or D)>>P(diagnosed B| D).

This is because D, S and B are ordered.

We may construct the surface by four kernels U_1, U_2, U_3, and U_4, assuming all 3 classes have equal probability to occur:

$$W_1 = 1 - Specificity_1 = \frac{1}{2}P(X_2 \le C_1) + \frac{1}{2}P(X_3 \le C_1) = E(U_1)$$

$$\text{with } U_1 = \frac{1}{2} I(X_2 \le C_1) + \frac{1}{2} I(X_3 \le C_1) \tag{2.2.5}$$

$$W_2 = 1 - Specificity_2 = P(X_3 \le C_2) = E(U_2) \qquad \text{with } U_2 = I(X_3 \le C_2) \tag{2.2.6}$$

$$W_3 = 1 - Sensitivity_1 = P(X_1 \ge C_1) = E(U_3) \qquad \text{with } U_3 = I(X_1 \ge C_1) \tag{2.2.7}$$

$$W_4 = 1 - Sensitivity_2 = \frac{1}{2} P(X_1 \ge C_2) + \frac{1}{2} P(X_2 \ge C_2) = E(U_4)$$

$$\text{with } U_4 = \frac{1}{2} I(X_1 \ge C_2) + \frac{1}{2} I(X_2 \ge C_2). \tag{2.2.8}$$

According to the representation of two cases above, we can define and express VUS as

$$VUS = E(\int \int\int\int_{C_1 \le C_2} dU_1 dU_2 dU_3 dU_4). \tag{2.2.9}$$

Here $\int \int\int\int_{C_1 \le C_2} dU_1 dU_2 dU_3 dU_4 \equiv \int_{C_2 \in (-\infty, +\infty)} \int_{C_1 \in (-\infty, C_2)} \int_{C_2 \in (C_1, +\infty)} \int_{C_1 \in (-\infty, C_2)} dU_1 dU_2 dU_3 dU_4$

(We can also define VUS directly by integration of W_i's but the estimate of it will be quite difficult. Therefore, we use mean of kernel integration to simplify it and make the inference possible.)

Theorem 1. With the set up as stated above and assuming equal probabilities from each of the three classes, the estimated VUS is

$$\hat{VUS} = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \frac{1}{4} (I(X_{2j} \le X_{3k}) + \frac{1}{2}) I(X_{3k} \ge X_{1i}) (I(X_{1i} \le X_{2j}) + \frac{1}{2})), \tag{2.2.10}$$

where $\{X_{ij} \mid j = 1,..., n_i\}$ and i=1, 2, 3 are three random samples from D, S, and B population, respectively.

Proof:

$$h(X_1, X_2, X_3) = \int \iiint_{C_1 \le C_2} dU_1 dU_2 dU_3 dU_4 = \int_{C_2 \in (-\infty, +\infty)} A_3 dU_4$$

where $A_3 = \int_{C_1 \in (-\infty, C_2)} A_2 dU_3$, $A_2 = \int_{C_2 \in (C_1, +\infty)} A_1 dU_2$, $A_1 = \int_{C_1 \in (-\infty, C_2)} dU_1$ (2.2.11)

$$A_1 = \int_{C_1 \in (-\infty, C_2)} dU_1 = \int_0^{\frac{1}{2}(I(X_2 \le C_2) + I(X_3 \le C_2))} dU_1 = \frac{1}{2}(I(X_2 \le C_2) + I(X_3 \le C_2)) \quad (2.2.12)$$

$$A_2 = \int_{C_2 \in (C_1, +\infty)} A_1 dU_2 = \frac{1}{2} \int_{I(X_3 \le C_1)}^{1} (I(X_2 \le C_2) + I(X_3 \le C_2)) dI(X_3 \le C_2)$$

$$= \frac{1}{2} I(X_3 \ge C_1)(I(X_2 \le X_3) + \frac{1}{2}) \quad (2.2.13)$$

$$A_3 = \int_{C_1 \in (-\infty, C_2)} A_2 dU_3 = \frac{1}{2}(I(X_2 \le X_3) + \frac{1}{2}) \int_0^{I(X_1 \le C_2)} I(X_3 \ge C_1) dI(X_1 \le C_1)$$

$$= \frac{1}{2}(I(X_2 \le X_3) + \frac{1}{2}) I(X_3 \ge X_1) I(X_1 \le C_2) \quad (2.2.14)$$

$$h(X_1, X_2, X_3) = \int \iiint_{C_1 \le C_2} dU_1 dU_2 dU_3 dU_4 = \int_{C_2 \in (-\infty, +\infty)} A_3 dU_4$$

$$= \frac{1}{4}(I(X_2 \le X_3) + \frac{1}{2}) I(X_3 \ge X_1) \int_0^1 I(X_1 \le C_2) d(I(X_1 \le C_2) + I(X_2 \le C_2))$$

$$= \frac{1}{4}(I(X_2 \le X_3) + \frac{1}{2}) I(X_3 \ge X_1)(I(X_1 \le X_2) + \frac{1}{2}) \quad (2.2.15)$$

Then,

$$VUS = E(\frac{1}{4}(I(X_2 \le X_3) + \frac{1}{2}) I(X_3 \ge X_1)(I(X_1 \le X_2) + \frac{1}{2})) \quad (2.2.16)$$

To estimate VUS:

$$\hat{VUS} = \hat{E}(\frac{1}{4}(I(X_2 \le X_3) + \frac{1}{2})I(X_3 \ge X_1)(I(X_1 \le X_2) + \frac{1}{2}))$$

$$= \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} h(X_{1i}; X_{2j}; X_{3k})$$

$$= \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \frac{1}{4}(I(X_{2j} \le X_{3k}) + \frac{1}{2})I(X_{3k} \ge X_{1i})(I(X_{1i} \le X_{2j}) + \frac{1}{2})) \quad (2.2.17)$$

Theorem 2. Let a_1, a_2, a_3 be the probabilities in D, S, and B, respectively. Then for $a_1 + a_2 + a_3 = 1$, we have the estimated VUS as

$$\hat{VUS} = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} (\frac{a_2}{a_2 + a_3} I(X_{2j} \le X_{3k}) + \frac{a_3}{2(a_2 + a_3)})(\frac{a_2}{a_1 + a_2} I(X_{1i} \le X_{2j}) + \frac{a_1}{2(a_1 + a_2)})I(X_{1i} \le X_{3k}) \cdot \quad (2.2.18)$$

Proof

Here we consider different probabilities from each of the three classes, then we need to re-define $U_i's$ and $W_i's$ as follow:

$$W_1 = 1 - Specificity_1 = \frac{a_2}{a_2 + a_3} P(X_2 \le C_1) + \frac{a_3}{a_2 + a_3} P(X_3 \le C_1) = E(U_1)$$

$$\text{with } U_1 = \frac{a_2}{a_2 + a_3} I(X_2 \le C_1) + \frac{a_3}{a_2 + a_3} I(X_3 \le C_1) \quad (2.2.19)$$

$$W_2 = 1 - Specificity_2 = P(X_3 \le C_2) = E(U_2) \qquad \text{with } U_2 = I(X_3 \le C_2) \quad (2.2.20)$$

$$W_3 = 1 - Sensitivity_1 = P(X_1 \ge C_1) = E(U_3) \qquad \text{with } U_3 = I(X_1 \ge C_1) \quad (2.2.21)$$

$$W_4 = 1 - Sensitivity_2 = \frac{a_1}{a_1 + a_2} P(X_1 \ge C_2) + \frac{a_2}{a_1 + a_2} P(X_2 \ge C_2) = E(U_4)$$

$$\text{with } U_4 = \frac{a_1}{a_1 + a_2} I(X_1 \geq C_2) + \frac{a_2}{a_1 + a_2} I(X_2 \geq C_2). \tag{2.2.22}$$

$$VUS = E(\int \iiint_{C_1 \leq C_2} dU_1 dU_2 dU_3 dU_4)$$

Here,

$$h(X_1, X_2, X_3) = \int \iiint_{C_1 \leq C_2} dU_1 dU_2 dU_3 dU_4 = \int_{C_2 \in (-\infty, +\infty)} A_3 dU_4$$

where $A_3 = \int_{C_1 \in (-\infty, C_2)} A_2 dU_3$, $A_2 = \int_{C_2 \in (C_1, +\infty)} A_1 dU_2$, $A_1 = \int_{C_1 \in (-\infty, C_2)} dU_1$

$$A_1 = \int_{C_1 \in (-\infty, C_2)} dU_1 = \int_0^{\frac{a_2}{a_2 + a_3} I(X_2 \leq C_2) + \frac{a_3}{a_2 + a_3} I(X_3 \leq C_2)} dU_1 = \frac{a_2}{a_2 + a_3} I(X_2 \leq C_2) + \frac{a_3}{a_2 + a_3} I(X_3 \leq C_2)$$

(2.2.23)

$$A_2 = \int_{C_2 \in (C_1, +\infty)} A_1 dU_2 = \int_{I(X_3 \leq C_1)}^1 (\frac{a_2}{a_2 + a_3} I(X_2 \leq C_2) + \frac{a_3}{a_2 + a_3} I(X_3 \leq C_2)) dI(X_3 \leq C_2)$$

$$= (\frac{a_2}{a_2 + a_3} I(X_2 \leq X_3) + \frac{a_3}{2(a_2 + a_3)}) I(X_3 \geq C_1) \tag{2.2.24}$$

$$A_3 = \int_{C_1 \in (-\infty, C_2)} A_2 dU_3 = (\frac{a_2}{a_2 + a_3} I(X_2 \leq X_3) + \frac{a_3}{2(a_2 + a_3)}) \int_0^{I(X_1 \leq C_2)} I(X_3 \geq C_1) dI(X_1 \leq C_1)$$

$$= (\frac{a_2}{a_2 + a_3} I(X_2 \leq X_3) + \frac{a_3}{2(a_2 + a_3)}) I(X_3 \geq X_1) I(X_1 \leq C_2) \tag{2.2.25}$$

$$h(X_1, X_2, X_3) = \int \iiint_{C_1 \leq C_2} dU_1 dU_2 dU_3 dU_4 = \int_{C_2 \in (-\infty, +\infty)} A_3 dU_4$$

$$= (\frac{a_2}{a_2 + a_3} I(X_2 \le X_3) + \frac{a_3}{2(a_2 + a_3)}) I(X_3 \ge X_1) \int_0^1 I(X_1 \le C_2)$$

$$d(\frac{a_1}{a_1 + a_2} I(X_1 \ge C_2) + \frac{a_2}{a_1 + a_2} I(X_2 \ge C_2))$$

$$= (\frac{a_2}{a_2 + a_3} I(X_2 \le X_3) + \frac{a_3}{2(a_2 + a_3)})(\frac{a_2}{a_1 + a_2} I(X_1 \le X_2) + \frac{a_1}{2(a_1 + a_2)}) I(X_1 \le X_3) \qquad (2.2.26)$$

Then,

$$VUS = E((\frac{a_2}{a_2 + a_3} I(X_2 \le X_3) + \frac{a_3}{2(a_2 + a_3)})(\frac{a_2}{a_1 + a_2} I(X_1 \le X_2) + \frac{a_1}{2(a_1 + a_2)}) I(X_1 \le X_3)) \qquad (2.2.27)$$

To estimate VUS:

$$\hat{VUS} = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} h(X_{1i}; X_{2j}; X_{3k})$$

$$= \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} (\frac{a_2}{a_2 + a_3} I(X_{2j} \le X_{3k}) + \frac{a_3}{2(a_2 + a_3)})(\frac{a_2}{a_1 + a_2} I(X_{1i} \le X_{2j}) + \frac{a_1}{2(a_1 + a_2)})$$

$$* I(X_{1i} \le X_{3k})$$

Corollary1. The new proposed VUS is the linear combination of AUC which includes suspicious class into disease class, AUC which includes suspicious class into benign class, AUC which exclude suspicious class and VUS proposed by Nakas & Yiannoutsos (2004).

Proof

From (2.2.27)

$$VUS = E((\frac{a_2}{a_2+a_3} I(X_2 \le X_3) + \frac{a_3}{2(a_2+a_3)})(\frac{a_2}{a_1+a_2} I(X_1 \le X_2) + \frac{a_1}{2(a_1+a_2)}) I(X_1 \le X_3))$$

$$= \frac{a_2^2}{(a_1+a_2)(a_2+a_3)} P(X_1 \le X_2 \le X_3) + \frac{a_2 a_3}{2(a_1+a_2)(a_2+a_3)} P(X_1 \le X_2, X_1 \le X_3)$$
$$+ \frac{a_1 a_2}{2(a_1+a_2)(a_2+a_3)} P(X_1 \le X_3, X_2 \le X_3) + \frac{a_1 a_3}{4(a_1+a_2)(a_2+a_3)} P(X_1 \le X_3)$$

$$= \frac{a_2^2}{(a_1+a_2)(a_2+a_3)} VUS_{NY} + \frac{a_2 a_3}{2(a_1+a_2)(a_2+a_3)} AUC_{2-3} + \frac{a_1 a_2}{2(a_1+a_2)(a_2+a_3)} AUC_{2-1}$$
$$+ \frac{a_1 a_3}{4(a_1+a_2)(a_2+a_3)} AUC_{-2},$$

where $VUS_{NY} = P(X_1 \le X_2 \le X_3)$ is the VUS proposed by Nakas & Yiannoutsos (2004), $AUC_{2-3} = P(X_1 \le X_2, X_1 \le X_3)$ is the AUC which includes suspicious class into benign class, $AUC_{2-1} = P(X_1 \le X_3, X_2 \le X_3)$ is the AUC which includes suspicious class into disease class, and $AUC_{-2} = P(X_1 \le X_3)$ is the AUC which excludes suspicious class.

Corollary2. The maximum and minimum values of VUS in Theorem 2 can be found as:

$$VUS_{Max} = \frac{(a_1/2 + a_2)(a_2 + a_3/2)}{(a_1+a_2)(a_2+a_3)}, \text{ and } VUS_{Min} = \frac{a_1 a_3/8 + a_2{}^2/6 + a_1 a_2/6 + a_2 a_3/6}{(a_1+a_2)(a_2+a_3)}.$$

(2.2.28)

The corollary can be shown when we replace $VUS_{NY}, AUC_{2-3}, AUC_{2-1}$ and AUC_{-2} by 1 in corollary1 for the maximum volume case. For minimum volume case, we can also prove it by treating probabilities of any order of $\{X_1, X_2, X_3\}$ to be 1/6 .

2.3 Performance and advantage of new proposed VUS

From corollary2 in 2.2 we can find that the new proposed VUS (denote by VUS_{NEW}) is the linear combination of four criteria VUS_{NY}, AUC_{2-3}, AUC_{2-1} and AUC_{-2} (Defined in Corollary 1 of the last section). If we consider unweighted case,

$$VUS_{NEW} = \frac{1}{4}VUS_{NY} + \frac{1}{8}AUC_{2-3} + \frac{1}{8}AUC_{2-1} + \frac{1}{16}AUC_{-2} \quad (2.3.1)$$

Here VUS_{NY} has highest weight because it included 3 classes and therefore loss less information. AUC_{2-3} and AUC_{2-1} have lower weight because it is calculated by including suspicious class into benign class or disease class and therefore loss more information than VUS_{NY}. AUC_{-2} has lowest weight because it is calculated by excluding suspicious class into consideration and therefore loss more information than AUC_{2-3} and AUC_{2-1}. Also you can adjust the weight by using weighted formula in corollary2 according to different consideration.

The new proposed VUS_{NEW} is clearly better than AUC_{2-3}, AUC_{2-1} and AUC_{-2} because it takes suspicious class into consideration. Also VUS_{NEW} is better than VUS_{NY} because in VUS_{NEW} we consider both sensitivity and specificity but in VUS_{NY} we only consider sensitivity. We can use a simple example to show the advantage of VUS_{NEW} to VUS_{NY}

Consider a three-level screening test index. It is screened to be benign if the value of index is between 0 and 1; It is screened to be suspicious if the value of index is between 1 and 1.5 and it is screened to have disease if the value of index is greater than 1.5. For simplicity, we consider 3 individuals to have the tests, one of which is in benign class (the test index is denoted by X_1), one of which is in suspicious class (the test index is denoted by X_2), one of which is in

disease class (the test index is denoted by X_3). Now we have 2 different tests. In test 1, $X_1 = 0.5, X_2 = 1.2, X_3 = 0$. In test 2, $X_1 = 0.5, X_2 = 1.2, X_3 = 1.1$.In both of this two tests, the benign and suspicious classes are corrected screened but disease class is not. The difference is that in test 1, the disease class is screened to benign and in test 2, the disease class is screened to suspicious. Therefore, it is clear that test 2 is better than test 1. Let's use VUS_{NEW} and VUS_{NY} to measure these two tests. It is easy to compute that $VUS_{NY}{}^{(1)} = VUS_{NY}{}^{(2)} = 0$, $VUS_{NEW}{}^{(1)} = 0, VUS_{NEW}{}^{(2)} = \frac{3}{16}$. Then, VUS_{NY} fails to measure the difference between test 1 and test 2 but VUS_{NEW} can.

The table below show the performance of VUS proposed by Nakas & Yiannoutsos (2004) and new proposed VUS for difference test result.

Table 2.3 Performance comparison for 1:1:1 population proportion

Test	VUS_{NY}	AUC_{2-3}	AUC_{2-1}	AUC_{-2}	VUS_{NEW}
$X_1 < X_2 < X_3$	1	1	1	1	$\frac{9}{16}$
$X_1 < X_3 < X_2$	0	1	0	1	$\frac{3}{16}$
$X_2 < X_1 < X_3$	0	0	1	1	$\frac{3}{16}$
$X_2 < X_3 < X_1$	0	0	0	0	0
$X_3 < X_1 < X_2$	0	0	0	0	0
$X_3 < X_2 < X_1$	0	0	0	0	0

Also it is available to adjust the value of new proposed VUS by adjusting the population proportion of the three classes. For instance, let the population proportion of class “benign”, “suspicious” and “disease” is 2: 1: 1, which means here are more “benign” cases than “suspicious” and “disease”. Then

$$VUS_{NEW} = \frac{1}{6}VUS_{NY} + \frac{1}{12}AUC_{2-3} + \frac{1}{6}AUC_{2-1} + \frac{1}{12}AUC_{-2} \tag{2.3.2}$$

Similar to above, we can see the performance of VUS proposed by Nakas & Yiannoutsos (2004) and new proposed VUS for difference test result (only 3 individuals in 3 classes separately):

Table 2.4 Performance comparison for 2:1:1 population proportion

Test	VUS_{NY}	AUC_{2-3}	AUC_{2-1}	AUC_{-2}	VUS_{NEW}
$X_1 < X_2 < X_3$	1	1	1	1	$\frac{1}{2}$
$X_1 < X_3 < X_2$	0	1	0	1	$\frac{1}{6}$
$X_2 < X_1 < X_3$	0	0	1	1	$\frac{1}{4}$
$X_2 < X_3 < X_1$	0	0	0	0	0
$X_3 < X_1 < X_2$	0	0	0	0	0
$X_3 < X_2 < X_1$	0	0	0	0	0

Apparently, test I make perfect decision, then the VUS value will be the highest value. Comparing test II and test III, test II reverse class 2 and 3 and correctly diagnose class 1; test III reverse class 1 and 2 and correctly diagnose class 3. Here VUS_{NEW} for test III performs better than test II because population size of class "benign" is larger, then in test II, the correctly diagnosis for class "benign" is by more chance.

2.4 A real example

In this section, we illustrate the application of the proposed new method for the sensitivity and specificity in 3 classes with the real ordered data. Cervical cancer is one of the most common cancers for women, and is also one of the easiest one to recover among all cancers, if it can be treated early. Thus, it is important to diagnose the cervical cancer early with accuracy. There are two main diagnostic tests, Pap smear and Colposcopy that are used in the diagnosis of Cervical Cancer. Current routine screening diagnosis is Pap smear, which can detect precancerous and cancerous cells on the cervix. If the Pap test has unclear result for abnormality, then the doctor may request to perform that test again. If Pap smear shows the significant abnormality, the Colposcopy that required the extensive training, experience and a significant effort in order to check for any abnormalities of the tissue surrounding the vagina and cervix may be requested to perform. However, these tests have many disadvantages such as excessive time consumption, economic inefficiency and low accuracy etc. In general, only 0.1% of about 10% of women who were identified to have abnormal tissue in the cervix in Pap smear proved to present abnormality. In addition, the abnormal and cancerous cells are missing in the half of Pap smear test result due to sampling errors. Thus, the earlier and more accurate detection for cervical cancer can help reduce the risk of death.

Spectrx, Inc. has developed a non-invasive method to diagnose cervical cancer. They used fluorescence and reflectance spectroscopy technology to produce a fiber optical and camera system. The system was introduced in 1999, and the hybrid device was developed from 2002 to 2006. There were 648 patients participated in the multicenter clinical trial. The device collected data from 56 spatial points on a circle surface of the cervix for each patient. The more detailed data collection and their variables can be seen in [5, 6]. The gold standard pathology yields the values 0, 1, 2, 2.5, 3, 3.2 and 3.5 for this response variable. The Food and Drug Administration (FDA) classifies {0, 1} as normal, {3, 3.2, 3.5} as cancer. However, {2, 2.5} can be classified as either normal or disease. With this ambiguous classification, Spectrx decided to delete these patients in the data set. Therefore, the problem becomes dichotomy, and the usual AUC can be easily used as the basis to select the model. We like to know if we treat {2} as the class between disease and normal, will we select the same model or not? In this case, our VUS can be applied as a three-class case. We used same models as previous studies (see [5]) but including all patients with gold standard value 2 as the suspicious class. We also used the same 10-fold cross validation techniques to estimate the shrinkages of each model. The VUSs were used to select the model. In many situations the selections are reverse the selection by AUCs. Since it became a four dimensional volume instead of two dimensional area, the maximum and minimum values of VUS are much smaller than that of the AUC. In order to make a more comparative numbers, we transformed VUS using both linear and quadratic formulae.

All models included not only the standardized original variables, but also many Boolean algebraic transformed variables. Since the variables are kind of lengthy, we do not list them here. However, they can be seen in [5, 6]. The comparison results are shown in the following table. We

may see we select model 1.03, while AUC select model 1.04 in the series model 1. We select model 2.12 instead model 2.11 selected by AUC in the series model 2.

Table 2.5 AUC and VUS for some models in cervical cancer diagnosis example

Model	Type	AUC	VUS	Mapping_VUS (linear)	Mapping_VUS (Quadratic)
1.03	Training	0.869336	0.294819	0.747506494	0.780501572
	10x validation	0.843569	0.272403	0.709680536	0.743149032
1.04	Training	0.861435	0.297352	0.751780985	0.784631895
	10x validation	0.840525	0.272853	0.710439835	0.743913632
2.0	training	0.763041	0.247273	0.667272926	0.699394165
	10x validation	0.762792	0.247196	0.667143249	0.699257042
2.1	training	0.759681	0.247144	0.667055008	0.699163720
	10x validation	0.758400	0.246780	0.666441140	0.698514238
2.11	training	0.792451	0.259172	0.687352900	0.720376263
	10x validation	0.789141	0.254386	0.679276543	0.711996201
2.12	training	0.796061	0.260739	0.689997149	0.723103106
	10x validation	0.775284	0.258856	0.686818805	0.719824485
2.13	training	0.796543	0.258079	0.685508837	0.718469718
	10x validation	0.784567	0.256242	0.682408718	0.715255439
2.14	training	0.791886	0.258559	0.686318313	0.719307117
	10x validation	0.780874	0.257897	0.685201403	0.718151475
2.15	training	0.792185	0.256649	0.683094709	0.715967681
	10x validation	0.781040	0.252838	0.676664713	0.709269313

CHAPTER 3

Inference of new VUS

3.1 Multi-sample U-statistics

The basic theory of U-statistics was developed by Hoeffding (1948a). Detailed expositions of the general topic may be found in Denker (1985) and Lee (1990). See also Fraser (1957) Chapter 6, Serfling (1980) Chapter 5, and Lehmann (1999) Chapter 6.

We say that $\theta(P)$ is an estimable parameter within $\mathcal{P}$ which is a family of probability measures on an arbitrary measurable space, if for some integer m there exists an unbiased estimator of $\theta(P)$ based on m i.i.d. random variables distributed according to P; that is, if there exists a real-valued measurable function $h(x_1,...,x_m)$ such that $E_P(h(X_1,...,X_m)) = \theta(P)$ for all P, when $X_1,...,X_m$ are i.i.d. with distribution P. The smallest integer m with this property is called the degree of $\theta(P)$.

For a real-valued measurable function, $h(x_1,...,x_m)$ and for a sample $X_1,...,X_n$, of size $n \geq m$ from a distribution P, a U-statistic with kernel h is defined as

$$U_n = U_n(h) = \frac{(n-m)!}{n!} \sum_{P_{m,n}} h(X_{i_1},...,X_{i_m}) \tag{3.1.1}$$

where the summation is over the set $P_{m,n}$ of all n!/(n-m)! permutations $(i_1,...,i_m)$ of size m chosen form (1,2,…,n). If the kernel, h, is symmetric in its arguments, U_n has the equivalent form

$$U_n = U_n(h) = \frac{1}{\binom{n}{m}} \sum_{C_{m,n}} h(X_{i_1},...,X_{i_m}) \tag{3.1.2}$$

Where the summation is over the set $C_{m,n}$ of all $\binom{n}{m}$ combinations of m intergers,

$i_1 < i_2 < ... < i_m$ chosen from (1,2,…,n).

And we have the following asymptotic convergence theorem:

$$\sqrt{n}(U_n - \theta) \xrightarrow{L} N(0, m^2\sigma^2) \tag{3.1.3}$$

Here $Var(U_n) \to m^2\sigma^2 / n$ for large n.

The important extension to k-sample problems for k>1 has been made by Lehmann (1951) [7]. The basic ideas are contained in the 3-sample case which is discussed here. Here $\mathcal{P}$ is a family of probability measures, (F, G, H).

Consider independent samples, $X_1,...,X_{n_1}$ from F(x), $Y_1,...,Y_{n_2}$ from G(y) and $Z_1,...,Z_{n_3}$ from H(z). Let $h(x_1,...,x_{m_1}, y_1,..., y_{m_2}, z_1,..., z_{m_3})$ be a kernel, and let $\mathcal{P}$ be the set of all vectors such that the expectation

$$\theta = \theta(F,G,H) = E_{F,G,H}\, h(X_1,..., X_{m_1}, Y_1,..., Y_{m_2}, Z_1,..., Z_{m_3}) \tag{3.1.4}$$

is finite. We may assume without loss of generality that h is symmetric under independent permutations of $x_1,...,x_{m_1}, y_1,..., y_{m_2}$ and $, z_1,..., z_{m_3}$. The corresponding U-statistic is

$$U_{n_1,n_2,n_3} = U(h) = \frac{1}{\binom{n_1}{m_1}\binom{n_2}{m_2}\binom{n_3}{m_3}} \sum h(X_{i_1},...,X_{i_{m_1}}, Y_{j_1},...,Y_{j_{m_2}}, Z_{k_1},...,Z_{k_{m_3}}) \tag{3.1.5}$$

where the sum is over all $\binom{n_1}{m_1}\binom{n_2}{m_2}\binom{n_3}{m_3}$ sets of subscripts such that

$1 \le i_1 < ... < i_{m_1} \le n_1, 1 \le j_1 < ... < j_{m_2} \le n_2$ and $1 \le k_1 < ... < k_{m_3} \le n_3$. Again, it is clear that U is an unbiased estimate of θ.

According to one sample U-statistic asymptotic theorem, we have the following. Let

$$\sigma_{ijk}^2 = Cov[h(X_1,...,X_i,X_{i+1},...,X_{m_1},Y_1,...,Y_j,Y_{j+1},...,Y_{m_2},Z_1,...,Z_k,Z_{k+1},...,Z_{m_3}),$$

$$h(X_1,...,X_i,X'_{i+1},...,X'_{m_1},Y_1,...,Y_j,Y'_{j+1},...,Y'_{m_2},Z_1,...,Z_k,Z'_{k+1},...,Z'_{m_3})] \tag{3.1,6}$$

where the X's, Y's and Z's are independently distributed according to F, G and H respectively.

For P$\in \mathcal{P}$,

$$Var(U_{n_1,n_2,n_3}) = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\sum_{k=1}^{m_3} \frac{\binom{n_1}{i}\binom{n_1-m_1}{m_1-i}}{\binom{n_1}{m_1}} \frac{\binom{n_2}{j}\binom{n_2-m_2}{m_2-j}}{\binom{n_2}{m_2}} \frac{\binom{n_3}{k}\binom{n_3-m_3}{m_3-k}}{\binom{n_3}{m_3}} \sigma_{ijk}^2 \tag{3.1.7}$$

Moreover, if σ_{m_1,m_2,m_3}^2 is finite, and if $n_1/N \to p_1 \in (0,1)$, $n_2/N \to p_2 \in (0,1)$ as $N = (n_1+n_2+n_3) \to \infty$, then

$$\sqrt{N}(U_{n_1,n_2,n_3} - \theta) \xrightarrow{L} N(0,\sigma^2), \tag{3.1.8}$$

where $\sigma^2 = \frac{m_1^2}{p_1}\sigma_{100}^2 + \frac{m_2^2}{p_2}\sigma_{010}^2 + \frac{m_3^2}{1-p_1-p_2}\sigma_{001}^2$

3.2 Inference of VUS using U-statistics

Apply the 3-sample U-statistics in section 3.1 to the VUS, we have the following theorem:

Theorem 3. Let U_{n_1,n_2,n_3} be the estimate of the VUS and VUS be the real value. n_1, n_2, n_3 are the sample size of each class. $n_1 / N \to a_1 \in (0,1)$, $n_2 / N \to a_2 \in (0,1)$ and $a_3 = 1 - a_1 - a_2$ as $N = (n_1 + n_2 + n_3) \to \infty$. Then when sample size is large enough,

$$\sqrt{N}(U_{n_1,n_2,n_3} - VUS) \xrightarrow{L} N(0, \sigma^2) \tag{3.2.1}$$

where, $\sigma^2 = \frac{1}{a_1}\sigma_{100}^2 + \frac{1}{a_2}\sigma_{010}^2 + \frac{1}{a_3}\sigma_{001}^2$

$$\sigma_{100}^2 = Cov(h(X_1; X_2; X_3), h(X_1; X_2'; X_3'))$$

$$\sigma_{010}^2 = Cov(h(X_1; X_2; X_3), h(X_1'; X_2; X_3'))$$

$$\sigma_{001}^2 = Cov(h(X_1; X_2; X_3), h(X_1'; X_2'; X_3))$$

$$h(X_1; X_2; X_3) = (\frac{a_2}{a_2 + a_3} I(X_2 \le X_3) + \frac{a_3}{2(a_2 + a_3)})(\frac{a_2}{a_1 + a_2} I(X_1 \le X_2) + \frac{a_1}{2(a_1 + a_2)}) I(X_1 \le X_3)$$

Lemma 1: Let T_N^*, N=1,2,… be a sequence of random variables, the distribution of which tend to a limit distribution L, and let T_N be another sequence satisfying $E(T_N^* - T_N)^2 \to 0$. Then, the distribution of T_N also trends to L.

Proof

Let $R_N = T_N^* - T_N$. Then $R_N \xrightarrow{P} 0$. Therefore, T_N and T_N^* have the same limit distribution.

Lemma 2: $Var\{\sqrt{N}(U_{n_1,n_2,n_3} - VUS)\} \to \sigma^2, N \to \infty$

Proof

$$Var(U_{n_1,n_2,n_3}) = E(U_{n_1,n_2,n_3}{}^2) - VUS^2 = (\frac{1}{n_1 n_2 n_3})^2 E((\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} h(X_{1i}; X_{2j}; X_{3k}))^2) - VUS^2$$

$$= (\frac{1}{n_1 n_2 n_3})^2 (\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\sum_{l=1}^{n_1}\sum_{m=1}^{n_2}\sum_{n=1}^{n_3} Eh(X_{1i}; X_{2j}; X_{3k}) h(X_{1l}; X_{2m}; X_{3n})) - VUS^2$$

$$= (\frac{1}{n_1 n_2 n_3})^2 (\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\sum_{l=1}^{n_1}\sum_{m=1}^{n_2}\sum_{n=1}^{n_3} [Cov(h(X_{1i}; X_{2j}; X_{3k}), h(X_{1l}; X_{2m}; X_{3n})) + VUS^2] - VUS^2 =$$

$$(\frac{1}{n_1 n_2 n_3})^2 (\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\sum_{l=1}^{n_1}\sum_{m=1}^{n_2}\sum_{n=1}^{n_3} Cov(h(X_{1i}; X_{2j}; X_{3k}), h(X_{1l}; X_{2m}; X_{3n}))$$

$$= (\frac{1}{n_1 n_2 n_3})^2 (n_1 n_2 n_3 Var(h(X_1; X_2; X_3)) + n_1(n_1 - 1) n_2 n_3 Cov(h(X_1; X_2; X_3), h(X_1'; X_2; X_3))$$

$$+ n_1 n_2 (n_2 - 1) n_3 Cov(h(X_1; X_2; X_3), h(X_1; X_2'; X_3))$$

$$+ n_1 n_2 n_3 (n_3 - 1) Cov(h(X_1; X_2; X_3), h(X_1; X_2; X_3'))$$

$$+ n_1(n_1 - 1) n_2 (n_2 - 1) n_3 Cov(h(X_1; X_2; X_3), h(X_1'; X_2'; X_3))$$

$$+ n_1(n_1 - 1) n_2 n_3 (n_3 - 1) Cov(h(X_1; X_2; X_3), h(X_1'; X_2; X_3'))$$

$$+ n_1 n_2 (n_2 - 1) n_3 (n_3 - 1) Cov(h(X_1; X_2; X_3), h(X_1; X_2'; X_3')))$$

$$= \frac{1}{n_1 n_2 n_3} Var(h(X_1; X_2; X_3)) + \frac{n_1 - 1}{n_1 n_2 n_3} Cov(h(X_1; X_2; X_3), h(X_1'; X_2; X_3))$$

$$+ \frac{n_2 - 1}{n_1 n_2 n_3} Cov(h(X_1; X_2; X_3), h(X_1; X_2'; X_3)) + \frac{n_3 - 1}{n_1 n_2 n_3} Cov(h(X_1; X_2; X_3), h(X_1; X_2; X_3'))$$

$$+ \frac{(n_1 - 1)(n_2 - 1)}{n_1 n_2 n_3} Cov(h(X_1; X_2; X_3), h(X_1'; X_2'; X_3))$$

$$+ \frac{(n_1 - 1)(n_3 - 1)}{n_1 n_2 n_3} Cov(h(X_1; X_2; X_3), h(X_1'; X_2; X_3'))$$

$$+ \frac{(n_2 - 1)(n_3 - 1)}{n_1 n_2 n_3} Cov(h(X_1; X_2; X_3), h(X_1; X_2'; X_3'))) \qquad (3.2.2)$$

Then

$$\sigma^2 = \lim_{N\to\infty} Var(\sqrt{n_1 + n_2 + n_3} U_{n_1, n_2, n_3}) = \lim_{N\to\infty} (n_1 + n_2 + n_3) Var(U_{n_1, n_2, n_3})$$

$$= \lim_{N\to\infty} [\frac{n_1 + n_2 + n_3}{n_1 n_2 n_3} Var(h(X_1; X_2; X_3)) + \frac{(n_1 - 1)(n_1 + n_2 + n_3)}{n_1 n_2 n_3} Cov(h(X_1; X_2; X_3), h(X_1'; X_2; X_3))$$

$$+ \frac{(n_2 - 1)(n_1 + n_2 + n_3)}{n_1 n_2 n_3} Cov(h(X_1; X_2; X_3), h(X_1; X_2'; X_3))$$

$$+ \frac{(n_3 - 1)(n_1 + n_2 + n_3)}{n_1 n_2 n_3} Cov(h(X_1; X_2; X_3), h(X_1; X_2; X_3'))$$

$$+\frac{(n_1-1)(n_2-1)(n_1+n_2+n_3)}{n_1n_2n_3}Cov(h(X_1;X_2;X_3),h(X_1';X_2';X_3))$$
$$+\frac{(n_1-1)(n_3-1)(n_1+n_2+n_3)}{n_1n_2n_3}Cov(h(X_1;X_2;X_3),h(X_1';X_2;X_3'))$$
$$+\frac{(n_2-1)(n_3-1)(n_1+n_2+n_3)}{n_1n_2n_3}Cov(h(X_1;X_2;X_3),h(X_1;X_2';X_3'))\ ]$$
$$=\frac{1}{p_1}Cov(h(X_1;X_2;X_3),h(X_1;X_2';X_3'))+\frac{1}{p_2}Cov(h(X_1;X_2;X_3),h(X_1';X_2;X_3'))$$

$$+\frac{1}{p_3}Cov(h(X_1;X_2;X_3),h(X_1';X_2';X_3)) \qquad (3.2.3)$$

This completes the proof of lemma 2.

Proof of theorem 3

Let $T_N=\sqrt{N}(U_{n_1,n_2,n_3}-VUS)$. By lemma 2, $Var(T_N)\to\sigma^2$.

Let

$$T_N^*=\sqrt{N}\{\frac{1}{n_1}\sum_{i=1}^{n_1}[\phi_{100}(X_{1i})-VUS]+\frac{1}{n_2}\sum_{j=1}^{n_2}[\phi_{010}(X_{2j})-VUS]+\frac{1}{n_3}\sum_{k=1}^{n_3}[\phi_{001}(X_{3k})-VUS]\} \quad (3.2.4)$$

where $\phi_{100}(x_1)=EU_{n_1,n_2,n_3}(x_1,X_2,X_3)$, $\phi_{010}(x_2)=EU_{n_1,n_2,n_3}(X_1,x_2,X_3)$,

$\phi_{001}(x_3)=EU_{n_1,n_2,n_3}(X_1,X_2,x_3)$

Then, $Var\phi_{100}(X_1)=\sigma_{100}^2$, $Var\phi_{010}(X_2)=\sigma_{010}^2$, $Var\phi_{001}(X_3)=\sigma_{001}^2$

Therefore, $VarT_N^*\to\sigma^2$

Notice, $Cov(T_N,T_N^*)=\frac{N}{n_1}Cov(U_{n_1,n_2,n_3},\sum_{i=1}^{n_1}\phi_{100}(X_{1i}))+\frac{N}{n_2}Cov(U_{n_1,n_2,n_3},\sum_{j=1}^{n_2}\phi_{010}(X_{2j}))$

$$+\frac{N}{n_3}Cov(U_{n_1,n_2,n_3},\sum_{k=1}^{n_3}\phi_{001}(X_{3k}))$$

$$=\sigma^2$$

Since $E(T_N^*) = E(T_N) = VUS$

$$E(T_N^* - T_N)^2 = Var(T_N^*) + Var(T_N) - 2Cov(T_N^*, T_N) = 0$$

By central limit theorem, $T_N^* \xrightarrow{L} N(0, \sigma^2)$

Then, by lemma 1, $T_N \xrightarrow{L} N(0, \sigma^2)$

This completes the proof of theorem 3.

When sample size is small, we use (3.2.2) to estimate σ^2:

$$\hat{\sigma}^2 = \frac{1}{n_1 n_2 n_3} \hat{Var}(h(X_1; X_2; X_3)) + \frac{n_1 - 1}{n_1 n_2 n_3} \hat{Cov}(h(X_1; X_2; X_3), h(X_1'; X_2; X_3))$$

$$+ \frac{n_2 - 1}{n_1 n_2 n_3} \hat{Cov}(h(X_1; X_2; X_3), h(X_1; X_2'; X_3)) + \frac{n_3 - 1}{n_1 n_2 n_3} \hat{Cov}(h(X_1; X_2; X_3), h(X_1; X_2; X_3'))$$

$$+ \frac{(n_1 - 1)(n_2 - 1)}{n_1 n_2 n_3} \hat{Cov}(h(X_1; X_2; X_3), h(X_1'; X_2'; X_3))$$

$$+ \frac{(n_1 - 1)(n_3 - 1)}{n_1 n_2 n_3} \hat{Cov}(h(X_1; X_2; X_3), h(X_1'; X_2; X_3'))$$

$$+ \frac{(n_2 - 1)(n_3 - 1)}{n_1 n_2 n_3} \hat{Cov}(h(X_1; X_2; X_3), h(X_1; X_2'; X_3'))) \qquad (3.2.5)$$

3.3 Empirical likelihood method

Empirical likelihood (EL) (Owen, 1990, 2001) [8] is a popular non-parametric method traditionally used for providing confidence intervals for means. The EL method has many advantages over other non-parametric methods. For example, it has better small sample performance than approaches based on normal approximation; empirical likelihood based confidence regions are range preserving and transformation respecting; the regularity conditions for empirical likelihood based methods are weak and natural. However, the empirical likelihood

method has not been used widely in the study of accuracy of diagnostic tests. Qin and Zhou (2005) [9] proposed empirical likelihood based confidence intervals for the area under the ROC curve. In this section, we are going to use the idea of empirical likelihood method to construct EL-based confidence intervals for the proposed VUS.

3.4 Inference of VUS using multi-sample empirical likelihood method

As the definition,

$$VUS = \theta = \frac{a_2^2}{(a_1+a_2)(a_2+a_3)} P(X_1 \le X_2 \le X_3) + \frac{a_2 a_3}{2(a_1+a_2)(a_2+a_3)} P(X_1 \le X_2, X_1 \le X_3)$$
$$+ \frac{a_1 a_2}{2(a_1+a_2)(a_2+a_3)} P(X_1 \le X_3, X_2 \le X_3) + \frac{a_1 a_3}{4(a_1+a_2)(a_2+a_3)} P(X_1 \le X_3)$$
$$h(X_1, X_2, X_3) = \frac{a_2^2}{(a_1+a_2)(a_2+a_3)} I(X_1 \le X_2 \le X_3) + \frac{a_2 a_3}{2(a_1+a_2)(a_2+a_3)} I(X_1 \le X_2, X_1 \le X_3)$$
$$+ \frac{a_1 a_2}{2(a_1+a_2)(a_2+a_3)} I(X_1 \le X_3, X_2 \le X_3) + \frac{a_1 a_3}{4(a_1+a_2)(a_2+a_3)} I(X_1 \le X_3)$$

$$\hat{\theta} = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} h(X_{1i}, X_{2j}, X_{3k}) \tag{3.4.1}$$

Notice $E(\hat{\theta} - \theta) = 0$. The empirical likelihood ratio can be expressed as

$$R(\theta) = \sup\{\prod_{i=1}^{3} \prod_{j=1}^{n_i} n_i u_{ij} \mid u_{ij} \ge 0, \sum_{j=1}^{n_i} u_{ij} = 1, (j = 1,..n_i, i = 1,2,3), \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} u_{1i} u_{2j} u_{3k} H_{ijk}(\theta) = 0\} \tag{3.4.2}$$

where $H_{ijk}(\theta) = h(X_{1i}, X_{2j}, X_{3k}) - \theta$

Then, the corresponding empirical log-likelihood ratio for VUS is $l(\theta) = -2\log R(\theta)$.

Theorem 4. If θ is the true value of VUS, then the limit distribution of empirical log-likelihood ratio $l(\theta)$ is a chi-square distribution with one degree of freedom. That is

$$l(\theta) \xrightarrow{L} \chi_1^2 \tag{3.4.3}$$

Proof.

$$R(\theta) = \sup\{\prod_{i=1}^{3}\prod_{j=1}^{n_i} n_i u_{ij} \mid u_{ij} \geq 0, \sum_{j=1}^{n_i} u_{ij} = 1, (j = 1,..n_i, i = 1,2,3), \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} u_{1i}u_{2j}u_{3k}H_{ijk}(\theta) = 0\}$$

Using Lagrange multiplier we find that

$$u_{1i} = \frac{1}{n_1 + \lambda\sum_{q=1}^{n_2}\sum_{r=1}^{n_3} u_{2q}u_{3r}H_{iqr}(\theta)} \quad i = 1,2,...,n_1 \tag{3.4.4}$$

$$u_{2j} = \frac{1}{n_2 + \lambda\sum_{p=1}^{n_1}\sum_{r=1}^{n_3} u_{1p}u_{3r}H_{pjr}(\theta)} \quad j = 1,2,...,n_2 \tag{3.4.5}$$

$$u_{3k} = \frac{1}{n_3 + \lambda\sum_{p=1}^{n_1}\sum_{q=1}^{n_2} u_{1p}u_{2q}H_{pqk}(\theta)} \quad k = 1,2,...,n_3 \tag{3.4.6}$$

where λ is defined by $\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} u_{1i}u_{2j}u_{3k}H_{ijk}(\theta) = 0$. (3.4.7)

We denote

$$\overline{H}_{i\bullet\bullet} = \frac{1}{n_2 n_3}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} H_{ijk}(\theta), \quad \overline{H}_{\bullet j\bullet} = \frac{1}{n_1 n_3}\sum_{i=1}^{n_1}\sum_{k=1}^{n_3} H_{ijk}(\theta), \quad \overline{H}_{\bullet\bullet k} = \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} H_{ijk}(\theta)$$

$$\tilde{H}_{i\bullet\bullet} = \sum_{q=1}^{n_2}\sum_{r=1}^{n_3} u_{2q}u_{3r}H_{iqr}(\theta), \quad \tilde{H}_{\bullet j\bullet} = \sum_{p=1}^{n_1}\sum_{r=1}^{n_3} u_{1p}u_{3r}H_{pjr}(\theta), \quad \tilde{H}_{\bullet\bullet k} = \sum_{p=1}^{n_1}\sum_{q=1}^{n_2} u_{1p}u_{2q}H_{pqk}(\theta)$$

and $\overline{H}_{\bullet\bullet\bullet} = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} H_{ijk}(\theta)$

Owen [8] have shown that $\lambda = o_p(1)$

Then, $u_{1i} = \frac{1}{n_1 + \lambda \tilde{H}_{i\bullet\bullet}} = \frac{1}{n_1}[1 - (\frac{\lambda}{n_1}\tilde{H}_{i\bullet\bullet}) + (\frac{\lambda}{n_1}\tilde{H}_{i\bullet\bullet})^2 - (\frac{\lambda}{n_1}\tilde{H}_{i\bullet\bullet})^3 + ...]$ (3.4.8)

$$u_{2j} = \frac{1}{n_2 + \lambda \tilde{H}_{\bullet j\bullet}} = \frac{1}{n_2}[1 - (\frac{\lambda}{n_2}\tilde{H}_{\bullet j\bullet}) + (\frac{\lambda}{n_2}\tilde{H}_{\bullet j\bullet})^2 - (\frac{\lambda}{n_2}\tilde{H}_{\bullet j\bullet})^3 + ...] \quad (3.4.9)$$

$$u_{3k} = \frac{1}{n_3 + \lambda \tilde{H}_{\bullet\bullet k}} = \frac{1}{n_3}[1 - (\frac{\lambda}{n_3}\tilde{H}_{\bullet\bullet k}) + (\frac{\lambda}{n_3}\tilde{H}_{\bullet\bullet k})^2 - (\frac{\lambda}{n_3}\tilde{H}_{\bullet\bullet k})^3 + ...] \quad (3.4.10)$$

Substitute them into $\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} u_{1i} u_{2j} u_{3k} H_{ijk}(\theta) = 0$

$$\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} [\frac{1}{n_1 n_2 n_3}(1 - \frac{\lambda}{n_1}\tilde{H}_{i\bullet\bullet} - \frac{\lambda}{n_2}\tilde{H}_{\bullet j\bullet} - \frac{\lambda}{n_3}\tilde{H}_{\bullet\bullet k})] H_{ijk}(\theta) \dot{=} 0 \quad (3.4.11)$$

Then, $\lambda \dot{=} D^{-1}\overline{H}_{\bullet\bullet\bullet}$ (3.4.12)

where $D = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} (\frac{1}{n_1}\tilde{H}_{i\bullet\bullet} + \frac{1}{n_2}\tilde{H}_{\bullet j\bullet} + \frac{1}{n_3}\tilde{H}_{\bullet\bullet k}) H_{ijk}(\theta)$

Since $\tilde{H}_{i\bullet\bullet} = \sum_{q=1}^{n_2} \sum_{r=1}^{n_3} u_{2q} u_{3r} H_{iqr}(\theta) \dot{=} \overline{H}_{i\bullet\bullet} - \frac{\lambda}{n_2 n_3} \sum_{q=1}^{n_2} \sum_{r=1}^{n_3} (\frac{1}{n_2}\tilde{H}_{\bullet q\bullet} + \frac{1}{n_3}\tilde{H}_{\bullet\bullet r}) H_{iqr}(\theta)$

$\tilde{H}_{i\bullet\bullet}$ can be replaced by $\overline{H}_{i\bullet\bullet}$ with the difference being absorbed into the coefficient of λ^2.

Similar to $\tilde{H}_{\bullet j\bullet}$ and $\tilde{H}_{\bullet\bullet k}$

Then, $D \dot{=} \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} (\frac{1}{n_1}\overline{H}_{i\bullet\bullet} + \frac{1}{n_2}\overline{H}_{\bullet j\bullet} + \frac{1}{n_3}\overline{H}_{\bullet\bullet k}) H_{ijk}(\theta)$

$$= \frac{1}{n_1^2} \sum_{i=1}^{n_1} \overline{H}_{i\bullet\bullet}^{\,2} + \frac{1}{n_2^2} \sum_{j=1}^{n_2} \overline{H}_{\bullet j\bullet}^{\,2} + \frac{1}{n_3^2} \sum_{k=1}^{n_3} \overline{H}_{\bullet\bullet k}^{\,2} \quad (3.4.13)$$

$$-2\log R(\theta) = 2\sum_{i=1}^{n_1}\log(1+\frac{\lambda}{n_1}\tilde{H}_{i\bullet\bullet}) + 2\sum_{j=1}^{n_2}\log(1+\frac{\lambda}{n_2}\tilde{H}_{\bullet j\bullet}) + 2\sum_{k=1}^{n_3}\log(1+\frac{\lambda}{n_3}\tilde{H}_{\bullet\bullet k})$$

$$\overset{\bullet}{=} 2\sum_{i=1}^{n_1}[\frac{\lambda}{n_1}\tilde{H}_{i\bullet\bullet} - \frac{1}{2}(\frac{\lambda}{n_1}\tilde{H}_{i\bullet\bullet})^2] + 2\sum_{j=1}^{n_2}[\frac{\lambda}{n_3}\tilde{H}_{\bullet j\bullet} - \frac{1}{2}(\frac{\lambda}{n_2}\tilde{H}_{\bullet j\bullet})^2]$$
$$+ 2\sum_{k=1}^{n_3}[\frac{\lambda}{n_3}\tilde{H}_{\bullet\bullet k} - \frac{1}{2}(\frac{\lambda}{n_3}\tilde{H}_{\bullet\bullet k})^2]$$

$$\overset{\bullet}{=} 2\sum_{i=1}^{n_1}[\frac{\lambda}{n_1}(\overline{H}_{i\bullet\bullet} - \frac{\lambda}{n_2 n_3}\sum_{q=1}^{n_2}\sum_{r=1}^{n_3}(\frac{1}{n_2}\overline{H}_{\bullet q\bullet} + \frac{1}{n_3}\overline{H}_{\bullet\bullet r})H_{iqr}(\theta)) - \frac{1}{2}(\frac{\lambda}{n_1}\overline{H}_{i\bullet\bullet})^2]$$
$$+ 2\sum_{j=1}^{n_2}[\frac{\lambda}{n_2}(\overline{H}_{\bullet j\bullet} - \frac{\lambda}{n_1 n_3}\sum_{p=1}^{n_1}\sum_{r=1}^{n_3}(\frac{1}{n_1}\overline{H}_{p\bullet\bullet} + \frac{1}{n_3}\overline{H}_{\bullet\bullet r})H_{pjr}(\theta)) - \frac{1}{2}(\frac{\lambda}{n_3}\overline{H}_{\bullet j\bullet})^2]$$
$$+ 2\sum_{k=1}^{n_3}[\frac{\lambda}{n_3}(\overline{H}_{\bullet\bullet k} - \frac{\lambda}{n_1 n_2}\sum_{p=1}^{n_1}\sum_{q=1}^{n_2}(\frac{1}{n_1}\overline{H}_{p\bullet\bullet} + \frac{1}{n_2}\overline{H}_{\bullet q\bullet})H_{pqk}(\theta)) - \frac{1}{2}(\frac{\lambda}{n_3}\overline{H}_{\bullet\bullet k})^2]$$

$$= 2\lambda\overline{H}_{\bullet\bullet\bullet} - \frac{2\lambda^2}{n_2^{\,2}}\sum_{q=1}^{n_2}\overline{H}_{\bullet q\bullet}^{\,2} - \frac{2\lambda^2}{n_3^{\,2}}\sum_{r=1}^{n_3}\overline{H}_{\bullet\bullet r}^{\,2} - \frac{\lambda^2}{n_1^2}\sum_{p=1}^{n_1}\overline{H}_{p\bullet\bullet}^{\,2}$$
$$+ 2\lambda\overline{H}_{\bullet\bullet\bullet} - \frac{2\lambda^2}{n_1^{\,2}}\sum_{p=1}^{n_1}\overline{H}_{p\bullet\bullet}^{\,2} - \frac{2\lambda^2}{n_3^{\,2}}\sum_{r=1}^{n_3}\overline{H}_{\bullet\bullet r}^{\,2} - \frac{\lambda^2}{n_2^2}\sum_{q=1}^{n_3}\overline{H}_{\bullet q\bullet}^{\,2}$$
$$+ 2\lambda\overline{H}_{\bullet\bullet\bullet} - \frac{2\lambda^2}{n_1^{\,2}}\sum_{p=1}^{n_1}\overline{H}_{p\bullet\bullet}^{\,2} - \frac{2\lambda^2}{n_2^{\,2}}\sum_{q=1}^{n_2}\overline{H}_{\bullet q\bullet}^{\,2} - \frac{\lambda^2}{n_3^2}\sum_{r=1}^{n_3}\overline{H}_{\bullet\bullet r}^{\,2}$$

$$= 6\lambda\overline{H}_{\bullet\bullet\bullet} - 5\lambda^2(\frac{1}{n_1^{\,2}}\sum_{p=1}^{n_1}\overline{H}_{p\bullet\bullet}^{\,2} + \frac{1}{n_2^{\,2}}\sum_{q=1}^{n_2}\overline{H}_{\bullet q\bullet}^{\,2} + \frac{1}{n_3^2}\sum_{r=1}^{n_3}\overline{H}_{\bullet\bullet r}^{\,2})$$

$$= 6\lambda\overline{H}_{\bullet\bullet\bullet} - 5\lambda^2 D$$

$$= \overline{H}_{\bullet\bullet\bullet}^{\,2} D^{-1} \qquad (3.4.14)$$

From (3.4.13), $D \overset{\bullet}{=} \frac{1}{n_1^{\,2}}\sum_{i=1}^{n_1}\overline{H}_{i\bullet\bullet}^{\,2} + \frac{1}{n_2^{\,2}}\sum_{j=1}^{n_2}\overline{H}_{\bullet j\bullet}^{\,2} + \frac{1}{n_3^2}\sum_{k=1}^{n_3}\overline{H}_{\bullet\bullet k}^{\,2}$

$$= \frac{1}{(n_1 n_2 n_3)^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\sum_{l=1}^{n_2}\sum_{m=1}^{n_3}H_{ijk}(\theta)H_{ilm}(\theta) + \frac{1}{(n_1 n_2 n_3)^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\sum_{l=1}^{n_1}\sum_{m=1}^{n_3}H_{ijk}(\theta)H_{ljm}(\theta)$$

$$+ \frac{1}{(n_1 n_2 n_3)^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\sum_{l=1}^{n_1}\sum_{m=1}^{n_2}H_{ijk}(\theta)H_{lmk}(\theta)$$

$$\frac{1}{n_1 n_2^2 n_3^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\sum_{l=1}^{n_2}\sum_{m=1}^{n_3}H_{ijk}(\theta)H_{ilm}(\theta)$$

$$=\frac{1}{n_1 n_2^2 n_3^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\sum_{l=1}^{n_2}\sum_{m=1}^{n_3}(h(X_{1i};X_{2j};X_{3k})-\theta)(h(X_{1i};X_{2l};X_{3m})-\theta)$$

$$\to Cov(h(X_1,X_2,X_3),h(X_1,X_2',X_3'))$$

Similarly, $\frac{1}{n_1^2 n_2 n_3^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\sum_{l=1}^{n_1}\sum_{m=1}^{n_3}H_{ijk}(\theta)H_{ljm}(\theta)\to Cov(h(X_1,X_2,X_3),h(X_1',X_2,X_3'))$

and $\frac{1}{n_1^2 n_2^2 n_3}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\sum_{l=1}^{n_1}\sum_{m=1}^{n_2}H_{ijk}(\theta)H_{lmk}(\theta)\to Cov(h(X_1,X_2,X_3),h(X_1,X_2',X_3'))$

From (3.2.2) $Var(\overline{H}_{\bullet\bullet\bullet})D^{-1}\overset{\bullet}{=}\frac{1}{D}(\frac{(n_1-1)(n_2-1)}{n_1n_2n_3}Cov(h(X_1;X_2;X_3),h(X_1';X_2';X_3))$

$$+\frac{(n_1-1)(n_3-1)}{n_1n_2n_3}Cov(h(X_1;X_2;X_3),h(X_1';X_2;X_3'))+\frac{(n_2-1)(n_3-1)}{n_1n_2n_3}Cov(h(X_1;X_2;X_3),h(X_1;X_2';X_3')))$$

$\to 1$ as $n_{1,}n_2,n_3\to\infty$

$\overline{H}_{\bullet\bullet\bullet}$ is asymptotic normal, then, $-2\log R(\theta)\to\chi^2_{(1)}$ as $n_{1,}n_2,n_3\to\infty$

The proof of Theorem 4 is thus complete.

3.5 Empirical likelihood confidence interval construction

Notice that, in order to find the lower bound and upper bound of empirical likelihood based confidence interval for VUS, we have to solve $n_1+n_2+n_3+2$ nonlinear equations, they are,

$$u_{1i}=\frac{1}{n_1+\lambda\sum_{q=1}^{n_2}\sum_{r=1}^{n_3}u_{2q}u_{3r}H_{iqr}(\theta)}\quad i=1,2,...,n_1 \tag{3.4.4}$$

$$u_{2j} = \frac{1}{n_2 + \lambda \sum_{p=1}^{n_1} \sum_{r=1}^{n_3} u_{1p} u_{3r} H_{pjr}(\theta)} \qquad j = 1,2,...,n_2 \tag{3.4.5}$$

$$u_{3k} = \frac{1}{n_3 + \lambda \sum_{p=1}^{n_1} \sum_{q=1}^{n_2} u_{1p} u_{2q} H_{pqk}(\theta)} \qquad k = 1,2,...,n_3 \tag{3.4.6}$$

$$\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} u_{1i} u_{2j} u_{3k} H_{ijk}(\theta) = 0 . \tag{3.4.7}$$

$$2\sum_{i=1}^{n_1} \log(1 + \frac{\lambda}{n_1} \tilde{H}_{i\bullet\bullet}) + 2\sum_{j=1}^{n_2} \log(1 + \frac{\lambda}{n_2} \tilde{H}_{\bullet j\bullet}) + 2\sum_{k=1}^{n_3} \log(1 + \frac{\lambda}{n_3} \tilde{H}_{\bullet\bullet k}) = \chi^{2\ (1-\alpha)}_{(1)} \tag{3.5.15}$$

Now, we have $n_1 + n_2 + n_3 + 2$ parameters ($u_{11},...u_{1n_1}, u_{21},...,u_{2n_2}, u_{31},...,u_{3n_3}, \lambda, \theta$).

The lower and upper limit of EL confidence interval will be the two solutions of θ. In order to find the numerical solutions of this equation system, an algorithm is proposed here:

Step I: Given initial value of $u_1 = (\frac{1}{n_1}, \frac{1}{n_1}, ..., \frac{1}{n_1})$, $u_2 = (\frac{1}{n_2}, \frac{1}{n_2}, ..., \frac{1}{n_2})$, $u_3 = (\frac{1}{n_3}, \frac{1}{n_3}, ..., \frac{1}{n_3})$.

Step II: Solve nonlinear equations (3.4.7) and (3.5.15) given initial $(\lambda_0, \theta_0) = (0, \theta_0)$.

Step III: Substitute the solutions of Step II into (3.4.4)-(3.4.6) to upgrade u_1, u_2 and u_3. Then go to Step II to solve (3.4.7) and (3.5.15) using the solution of Step II as initial value until converge.

According to the EL theory, given large initial value of θ_0, the final solution of θ will converge to EL upper limit and given small initial value of θ_0, the final solution of θ will converge to EL lower limit.

CHAPTER 4

VUS for any number of classes

It is straightforward to generalize new proposed VUS to k-class case as follows: Let the k classes be "L_1", "L_2" … "L_k". Also, let X_i be the classifier in k populations with i=1, 2…k representing $L_1, L_2, ..., L_k$ respectively. If we use $C_1, C_2, ..., C_{k-1}$ as the cutoff points, then we have similar table as 3-class case:

Table 4.1 Multi-class diagnosis table

		Actual Diseased Status			
		Level-1	Level-2	… …	Level-k
Test	Level-1	h_1	F_{12}	… …	F_{1k}
	Level-2	F_{21}	h_2	… …	F_{2k}
	… …	… …	… …	… …	… …
	Level-k	F_{k1}	F_{k2}	… …	h_k

Similar to 3-class case, we have k-1 sensitivities and k-1 specificities, we denote

$$Sensitivity_i = P(diagnosed \le i \mid real \le i) \quad \text{i=1,2,…,k-1} \tag{4.1}$$

$$Specificity_i = P(diagnosed > i \mid real > i) \quad \text{i=1,2,…,k-1} \tag{4.2}$$

Assume the population ratio for the k-level is $a_1 : a_2 : a_3 : ... : a_k$ ($\sum_{i=1}^{k} a_i = 1$) then,

For specificity,

$$W_{11} = 1 - Specificity_1 = \frac{1}{a_2 + ... + a_k} \sum_{i=2}^{k} a_i P(X_i \le C_1) = E(U_{11}) \quad (4.3)$$

where $U_{11} = \frac{1}{a_2 + ... + a_k} \sum_{i=2}^{k} a_i I(X_i \le C_1)$

$$W_{12} = 1 - Specificity_2 = \frac{1}{a_3 + ... + a_k} \sum_{i=3}^{k} a_i P(X_i \le C_2) = E(U_{12}) \quad (4.4)$$

where $U_{12} = \frac{1}{a_3 + ... + a_k} \sum_{i=3}^{k} a_i I(X_i \le C_2)$

…

$$W_{1p} = 1 - Specificity_p = \frac{1}{a_{p+1} + ... + a_k} \sum_{i=p+1}^{k} a_i P(X_i \le C_p) = E(U_{1p}) \quad (4.5)$$

where $U_{1p} = \frac{1}{a_{p+1} + ... + a_k} \sum_{i=p+1}^{k} a_i I(X_i \le C_p)$

…

$$W_{1,k-1} = 1 - Specificity_{k-1} = P(X_k \le C_{k-1}) = E(U_{1,k-1}) \quad (4.6)$$

where $U_{1,k-1} = I(X_k \le C_{k-1})$

For sensitivity,

$$W_{21} = 1 - Sensitivity_1 = P(X_1 \ge C_1) = E(U_{21}) \quad (4.7)$$

where $U_{21} = I(X_1 \ge C_1)$

$$W_{22} = 1 - Sensitivity_2 = \frac{1}{a_1 + .a_2} \sum_{i=1}^{2} a_i P(X_i \ge C_2) = E(U_{22}) \quad (4.8)$$

$$\text{where } U_{22} = \frac{1}{a_1 + .a_2} \sum_{i=1}^{2} a_i I(X_i \geq C_2)$$

…

$$W_{2p} = 1 - Sensitivity_p = \frac{1}{a_1 + ... + a_p} \sum_{i=1}^{p} a_i P(X_i \geq C_p) = E(U_{2p}) \tag{4.9}$$

$$\text{where } U_{2p} = \frac{1}{a_1 + ... + a_p} \sum_{i=1}^{p} a_i I(X_i \geq C_p)$$

…

$$W_{2,k-1} = 1 - Sensitivity_{k-1} = \frac{1}{a_1 + ... + a_{k-1}} \sum_{i=1}^{k-1} a_i P(X_i \geq C_{k-1}) = E(U_{2,k-1}) \tag{4.10}$$

$$\text{where } U_{2,k-1} \frac{1}{a_1 + ... + a_{k-1}} \sum_{i=1}^{k-1} a_i I(X_i \geq C_{k-1})$$

Here, $U_{11},...U_{1,k-1},U_{21},...,U_{2,k-1}$ are the kernels of $W_{11},...W_{1,k-1},W_{21},...,W_{2,k-1}$ separately.

Similar to 3-class case, we can define and express VUS as

$$VUS_k = E(\int_{C_1 \leq C_2 \leq} ... \int_{...\leq C_{k-1}} dU_{11}dU_{12}...dU_{1,k-1}dU_{21}dU_{22}...dU_{2,k-1}) \tag{4.11}$$

Here,

$$\int_{C_1 \leq C_2 \leq} ... \int_{...\leq C_{k-1}} dU_{11}dU_{12}...dU_{1,k-1}dU_{21}dU_{22}...dU_{2,k-1}$$
$$\equiv . \int_{C_{k-1} \in (-\infty,+\infty)} \int_{C_{k-2} \in (C_{k-3},+\infty)} ... \int_{C_2 \in (C_1,+\infty)} \int_{C_{k-1} \in (C_1,+\infty)} .. \int_{C_2 \in (-\infty,C_3)} \int_{C_1 \in (-\infty,C_2)} dU_{11}dU_{12}...dU_{1,k-1}dU_{21}dU_{22}...dU_{2,k-1}$$

(4.12)

Theorem 4. With the set up as stated above, the estimated VUS is

$$\hat{VUS} = \frac{1}{n_1...n_k} \sum_{p_1=1}^{n_1} \sum_{p_2=1}^{n_2} ... \sum_{p_k=1}^{n_k} K(X_{1p_1}, X_{2p_2}, ..., X_{kp_k}), \tag{4.13}$$

where

$$K(X_1, X_2, ..., X_k) = [\prod_{i=2}^{k-1}(a_i + a_{i+1} + ... + a_k)^{-1}][\prod_{j=2}^{k-1}(a_1 + a_2 + ... + a_j)^{-1}]I(X_1 \le X_k)$$
$$\sum_{i_{k-2}=k-1}^{k} ... \sum_{i_2=3}^{k}\sum_{i_1=2}^{k}\sum_{j_1=1}^{2}\sum_{j_2=1}^{3}... \sum_{j_{k-2}=1}^{k-1} a_{i_1}...a_{i_{k-2}}a_{j_1}...a_{j_{k-2}} I(X_{i_1} \le X_{i_2} ... \le X_k)I(X_1 \le ...X_{j_{k-3}} \le X_{j_{k-2}}) \tag{4.14}$$

and $\{X_{ij} \mid j = 1,..., n_i\}$, i=1, 2,…,k are k random samples from $L_1, L_2, ..., L_k$ population, respectively.

Proof,

The proof of this theorem is similar to 3-class case but the procedure is massy. We sketch the proof as follows:.

$$A_{11} = \int_{C_1 \in (-\infty, C_2)} dU_{11} = \frac{1}{a_2 + ... + a_k}\sum_{i=2}^{k} a_i I(X_i \le C_2) \tag{4.15}$$

$$A_{12} = \int_{C_2 \in (-\infty, C_3)} A_{11} dU_{12} = \frac{1}{(a_3 + ... + a_k)}\frac{1}{(a_2 + ... + a_k)} \int_{C_2 \in (-\infty, C_3)} \sum_{i=2}^{k} a_i I(X_i \le C_2) d\sum_{j=3}^{k} a_j I(X_j \le C_2)$$

$$= \frac{1}{(a_3 + ... + a_k)}\frac{1}{(a_2 + ... + a_k)}\sum_{i=2}^{k}\sum_{j=3}^{k} a_i a_j I(X_i \le X_j)(\sum_{j=3}^{k} I(X_j \le C_3)) \tag{4.16}$$

…

$$A_{1,k-2} = \int_{C_{k-2} \in (-\infty, C_{k-1})} A_{1,k-3} dU_{1,k-2}$$
$$= \prod_{j=2}^{k-1}\frac{1}{a_j + a_{j+1} + ... + a_k}\sum_{i_1=2}^{k}\sum_{i_2=3}^{k}... \sum_{i_{k-2}=k-1}^{k} a_{i_1}a_{i_2}...a_{i_{k-2}} I(X_{i_1} \le X_{i_2} ... \le X_{i_{k-2}})(\sum_{i_{k-2}=k-1}^{k} I(X_{i_{k-2}} \le C_{k-1})) \tag{4.17}$$

$$A_{20} = \int_{C_{k-1}\in(C_1,+\infty)} A_{1,k-2}\,dU_{1,k-1}$$

$$= \prod_{j=2}^{k-1} \frac{1}{a_j + a_{j+1} + ... + a_k} \sum_{i_1=2}^{k}\sum_{i_2=3}^{k} ... \sum_{i_{k-2}=k-1}^{k} a_{i_1} a_{i_2} ... a_{i_{k-2}} I(X_{i_1} \le X_{i_2} ... \le X_{i_{k-2}})$$

$$* (\int_{C_{k-1}\in(C_1,+\infty)} \sum_{i_{k-2}=k-1}^{k} I(X_{i_{k-2}} \le C_{k-1}) dI(X_k \le C_{k-1}))$$

$$= \prod_{j=2}^{k-1} \frac{1}{a_j + a_{j+1} + ... + a_k} \sum_{i_1=2}^{k}\sum_{i_2=3}^{k} ... \sum_{i_{k-2}=k-1}^{k} a_{i_1} a_{i_2} ... a_{i_{k-2}} I(X_{i_1} \le X_{i_2} ... \le X_{i_{k-2}} \le X_k) I(X_k \ge C_1) \qquad (4.18)$$

$$A_{21} = \int_{C_1\in(C_2,+\infty)} A_{20}\,dU_{21}$$

$$= \prod_{j=2}^{k-1} \frac{1}{a_j + a_{j+1} + ... + a_k} \sum_{i_1=2}^{k}\sum_{i_2=3}^{k} ... \sum_{i_{k-2}=k-1}^{k} a_{i_1} a_{i_2} ... a_{i_{k-2}} I(X_{i_1} \le X_{i_2} ... \le X_{i_{k-2}} \le X_k)$$

$$* \int_{C_1\in(C_2,+\infty)} I(X_k \ge C_1) dI(X_1 \ge C_1)$$

$$= \prod_{j=2}^{k-1} \frac{1}{a_j + a_{j+1} + ... + a_k} \sum_{i_1=2}^{k}\sum_{i_2=3}^{k} ... \sum_{i_{k-2}=k-1}^{k} a_{i_1} a_{i_2} ... a_{i_{k-2}} I(X_{i_1} \le X_{i_2} ... \le X_{i_{k-2}} \le X_k) I(X_k \ge X_1) I(X_1 \ge C_2)$$

(4.19)

…

$$A_{2,k-1} = \int_{C_{k-1}\in(-\infty,+\infty)} A_{2,k-2}\,dU_{2,k-1} = [\prod_{i=2}^{k-1} (a_i + a_{i+1} + ... + a_k)^{-1}][\prod_{j=2}^{k-1} (a_1 + a_2 + ... + a_j)^{-1}] I(X_1 \le X_k)$$

$$\sum_{i_{k-2}=k-1}^{k} ... \sum_{i_2=3}^{k}\sum_{i_1=2}^{k}\sum_{j_1=1}^{2}\sum_{j_2=1}^{3} ... \sum_{j_{k-2}=1}^{k-1} a_{i_1} ... a_{i_{k-2}} a_{j_1} ... a_{j_{k-2}} I(X_{i_1} \le X_{i_2} ... \le X_k) I(X_1 \le ... X_{j_{k-3}} \le X_{j_{k-2}})$$

(4.20)

Denote $A_{2,k-1}$ by $K(X_1, X_2, ..., X_k)$. Then,

$$VUS = E(A_{2,k-1}) = E(K(X_1, X_2, ..., X_k)).$$

$$\hat{VUS} = \hat{E}(K(X_1, X_2, ..., X_k)) = \frac{1}{n_1 ... n_k} \sum_{p_1=1}^{n_1}\sum_{p_2=1}^{n_2} ... \sum_{p_k=1}^{n_k} K(X_{1p_1}, X_{2p_2}, ..., X_{kp_k})) \qquad (4.21)$$

When $i_j = i_k$ then we have $I(X_{i_j} \le X_{i_k}) = \frac{1}{2}$

and, when $i_j = i_k = i_l$ then we have $I(X_{i_j} \le X_{i_k} \le X_{i_l}) = \frac{1}{6}$

and so on.

Here when k=3,

$$K(X_1, X_2, X_3) = (a_1 + a_2)^{-1}(a_2 + a_3)^{-1} I(X_1 \le X_3) \sum_{i=2}^{3} \sum_{j=1}^{2} a_i a_j I(X_i \le X_3) I(X_1 \le X_j)$$

$$= \frac{1}{(a_1 + a_2)(a_2 + a_3)} I(X_1 \le X_3)[a_1 a_2 I(X_1 \le X_1) I(X_2 \le X_3) + a_1 a_3 I(X_1 \le X_1) I(X_3 \le X_3)$$

$$+ {a_2}^2 I(X_1 \le X_2) I(X_2 \le X_3) + a_2 a_3 I(X_1 \le X_2) I(X_3 \le X_3)]$$

$$= \frac{1}{(a_1 + a_2)(a_2 + a_3)} I(X_1 \le X_3)[\frac{1}{2} a_1 a_2 I(X_2 \le X_3) + \frac{1}{4} a_1 a_3 + {a_2}^2 I(X_1 \le X_2 \le X_3)$$
$$+ \frac{1}{2} a_2 a_3 I(X_1 \le X_2)]$$

This is the same as we find in chapter 2.

When k=4,

$$K(X_1, X_2, X_3, X_4) = \frac{I(X_1 \le X_4) \sum_{i=3}^{4} \sum_{j=2}^{4} \sum_{k=1}^{2} \sum_{l=1}^{3} a_i a_j a_k a_l I(X_j \le X_i \le X_4) I(X_1 \le X_k \le X_l)}{(a_1 + a_2)(a_3 + a_4)(a_1 + a_2 + a_3)(a_2 + a_3 + a_4)}$$

if $a_1 = a_2 = a_3 = a_4$,

$$K(X_1, X_2, X_3, X_4) = \frac{1}{36} I(X_1 \le X_4)[3I(X_1 \le X_2 \le X_3 \le X_4) + I(X_1 \le X_2) I(X_3 \le X_4)$$
$$+ \frac{1}{2}(I(X_1 \le X_3 \le X_4) + I(X_1 \le X_2 \le X_4) + I(X_1 \le X_2 \le X_3, X_4) + I(X_1, X_2 \le X_3 \le X_4))$$
$$+ \frac{1}{3}(I(X_1 \le X_2) + I(X_3 \le X_4) + I(X_1 \le X_2 \le X_3) + I(X_2 \le X_3 \le X_4))$$
$$+ \frac{1}{4} I(X_1 \le X_3) I(X_2 \le X_4) + \frac{1}{6}(I(X_1 \le X_3) + I(X_2 \le X_4)) + \frac{1}{9}]$$

U-statistics is also available to do inference, but the calculation of asymptotic variance is massy when the number of class levels is large. So we skip this part in this dissertation.

CHAPTER 5

Simulation study

5.1 Calculation for theoretical VUS value

From corollary2 in chapter2 we have derived

$$VUS_{NEW} = \frac{a_2^2}{(a_1+a_2)(a_2+a_3)} VUS_{NY} + \frac{a_2 a_3}{2(a_1+a_2)(a_2+a_3)} AUC_{2-3}$$

$$+ \frac{a_1 a_2}{2(a_1+a_2)(a_2+a_3)} AUC_{2=1} + \frac{a_1 a_3}{4(a_1+a_2)(a_2+a_3)} AUC_{-2}, \qquad (5.1.1)$$

where $VUS_{NY} = P(X_1 \le X_2 \le X_3)$, $AUC_{2-3} = P(X_1 \le X_2, X_3)$,

$AUC_{2-1} = P(X_1, X_2 \le X_3)$ and $AUC_{-2} = P(X_1 \le X_3)$

If $F_i(x)$ is cumulative distribution of X_i and $f_i(x)$ is its density for i=1,2,3. Then

$$AUC_{2-1} = P(X_1, X_2 \le X_3) = \iiint_{x,y \le z} f_1(x) f_2(y) f_3(z) dxdydz = \int_{-\infty}^{+\infty} F_1(x) F_2(x) f_3(x) dx \qquad (5.1.2)$$

$$AUC_{2-3} = P(X_1 \le X_2, X_3) = \iiint_{x \le y,z} f_1(x) f_2(y) f_3(z) dxdydz = \int_{-\infty}^{+\infty} f_1(x)(1-F_2(x))(1-F_3(x)) dx$$

(5.1.3)

$$VUS_{NY} = P(X_1 \le X_2 \le X_3) = \iiint_{x \le y \le z} f_1(x) f_2(y) f_3(z) dxdydz = \int_{-\infty}^{+\infty} F_1(x) f_2(x)(1-F_3(x)) dx$$

(5.1.4)

$$AUC_{-2} = P(X_1 \le X_3) = \iint_{x \le z} f_1(x) f_3(z) dxdz = \int_{-\infty}^{+\infty} f_1(x)(1-F_3(x)) dx \qquad (5.1.5)$$

Gaussian Quadrature method is used to calculate numerical integration for our simulation study.

In numerical analysis, a quadrature rule is an approximation of the definite integral of function, usually stated as a weighted sum of function values at specified points within the domain of integration. An n-point Gaussian quadrature rule, named after Carl Friedrich Gauss, is a quadrature rule constructed to yield an exact result for polynomials of degree 2n-1 or less by a suitable choice of the points x_i and weights w_i for i=1,…,n. The domain of integration for such a rule is conventionally take as [-1, 1], so the rule is stated as

$$\int_{-1}^{1} f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i) \tag{5.1.6}$$

It can be shown (see Press, et al., or Stoer and Bulirsch) that the evaluation points are just the roots of a polynomial belonging to a class of orthogonal polynomials.

For the integration problem stated above, the associated polynomials are Legendre polynomials, $P_n(x)$. With the n^{th} polynomial normalized to give $P_n(x) = 1$, the i^{th} Gauss node, x_i, is the i^{th} root of $P_n(x)$, its weight is given by (Abramowitz & Stegun 1972, p. 887)

$$w_i = \frac{2}{(1-x_i^2)(P'_n(x_i))^2} \tag{5.1.8}$$

Some low-order rules for solving the integration problem are listed in Table 5.1.

An integral over [a, b] must be changed into an integral over [-1,1] before applying the Gaussian quadrature rule. This change of interval can be done in the following way:

$$\int_a^b f(x)dx = \frac{b-a}{2}\int_0^1 f(\frac{b-a}{2}x + \frac{a+b}{2})dx \tag{5.1.9}$$

Table 5.1 Gaussian quadrature points and weights

Number of points, n	Points, x_i	Weights w_i
1	0	2
2	$\pm\sqrt{1/3}$	1
3	0	8/9
	$\pm\sqrt{3/5}$	5/9
4	$\pm\sqrt{3-2\sqrt{6/5}/7}$	$\frac{18+\sqrt{30}}{36}$
	$\pm\sqrt{3+2\sqrt{6/5}/7}$	$\frac{18-\sqrt{30}}{36}$
5	0	128/225
	$\pm\frac{1}{3}\sqrt{5-2\sqrt{10/7}}$	$\frac{322+13\sqrt{70}}{900}$
	$\pm\frac{1}{3}\sqrt{5+2\sqrt{10/7}}$	$\frac{322-13\sqrt{70}}{900}$

An integral over [a, b] must be changed into an integral over [-1,1] before applying the Gaussian quadrature rule. This change of interval can be done in the following way:

$$\int_a^b f(x)dx = \frac{b-a}{2}\int_0^1 f(\frac{b-a}{2}x+\frac{a+b}{2})dx \tag{5.1.9}$$

After applying the Gaussian quadrature rule, the following approximation is obtained:

$$\frac{b-a}{2}\sum_{i=1}^{n} w_i f(\frac{b-a}{2}x_i+\frac{a+b}{2}) \tag{5.1.10}$$

Gaussian quadrature method is used to calculate definite integral in limited domain. But the integral domain we try to calculate is $(-\infty,+\infty)$. However, the functions we integrate converge to 0 when x goes to $-\infty$ and $+\infty$. Then, when we take small enough lower limit and large enough upper limit, the integral will be close to true value.

In order to make the theoretical value of VUS more accurate, we select five points for the Gaussian quadrature in this dissertation.

5.2 VUS mapping and theoretical value selection.

In this section, we try to compare the performance of different inference methods for different VUS values. But the maximum and minimum values of VUS are much smaller than that of the AUC because they are measures from different dimension space. In order to make a comparative numbers, proper transformation from VUS to AUC is necessary. Here we propose two mapping methods.

The most straightforward mapping is linear mapping. The formula is as follow:

$$\frac{VUS - VUS_{Min}}{VUS_{Max} - VUS_{Min}} = \frac{AUC - AUC_{Min}}{AUC_{Max} - AUC_{Min}} \tag{5.2.1}$$

Here the maximum and minimum value of AUC is 1 and 0.5, and the maximum, and the minimum values of VUS are derived in corollary 2 in chapter 2.

When d=3, VUS is the volume from 2(d-1)=4 dimensional space, and AUC is from a two dimensional space. Hence, it is more reasonable to modify linear mapping to quadratic mapping for d=3. Since the unit for dimension 4 is square of the dimension 2, quadratic transformation might be more reasonable. Hence, before mapping VUS to AUC, taking square root to VUS can

balance the dimension difference between VUS and AUC. Then the transformation formula is as follow:

$$\frac{\sqrt{VUS}-\sqrt{VUS_{Min}}}{\sqrt{VUS_{Max}}-\sqrt{VUS_{Min}}}=\frac{AUC-AUC_{Min}}{AUC_{Max}-AUC_{Min}} \tag{5.2.2}$$

We can also illustrate the above mapping methods by the graph below:

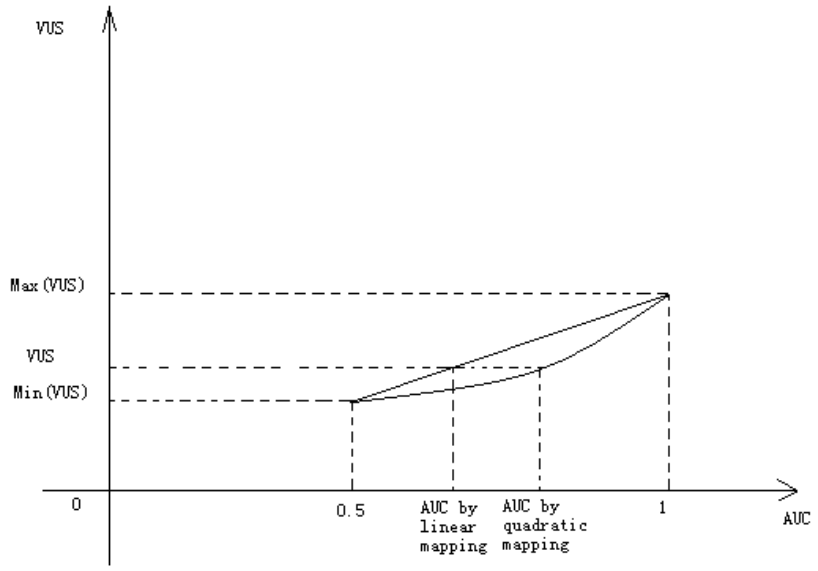

Figure 5.1 VUS transformation

In this simulation studies, we properly select the distribution of three classes such that the value of VUS is “small”, “median” and “large” separately. Since it is difficult to get the fix VUS by selection the parameters of distribution, we only choose distributions such that the AUC mapping from small VUS is around 0.65, the AUC mapping from median VUS is around 0.75 and the AUC mapping from large VUS is around 0.9. For each level of VUS, two kinds of locations are selected. For normal case, one kind is the mean of “suspicious” is in the middle of

"Disease" and "Benign". And another kind is mean of "suspicious" is bias to "Benign". For gamma case, one kind is the beta of "suspicious" is in the middle of "Disease" and "Benign". And another kind is beta of "suspicious" is bias to "Benign".

The population proportion we select in this simulation studies is 2:1:2.

Our simulation studies cover the cases of normal distribution and gamma distribution and the VUS values table and graphs are list in Table 5.2-Table 5.3 and Figure 5.2- Figure 5.3.

5.3 Simulation method and conclusion

Our simulation studies cover the cases of normal distribution and gamma distribution for X_1, X_2, X_3, with sample sizes 40 (small sample) and 100 (large sample). In total, 300 confidence intervals were constructed for each simulation. For normal distribution, we simulated with various values of (μ_1, μ_2, μ_3) and $(\sigma_1, \sigma_2, \sigma_3)$. Similarly, various values of parameters for gamma distribution are used. Three methods of constructing confidence intervals are used for comparison purpose. First two are the proposed U-statistic method and empirical likelihood method, while the other one is the bootstrap method. Using 95% confidence interval, we calculate the coverage rate of the theoretical VUS, which is computed using Gaussian quadrature to do the integrations. From Table 5.4 and Table 5.5, both normal and gamma distributions behave about the same in terms of the coverage rate. Except for the high volume and small sample case, the U-statistics method seems outperform the bootstrap case for the coverage rate. This indicates that the method of U-statistic indeed has its merit. The empirical likelihood has significantly shorter interval length than U-statistic and Bootstrap method. However, they always seems to have the least coverage probabilities.

Table 5.2 Theoretical VUS for normal distribution

$a_1 : a_2 : a_3 = 2:1:2$	(μ_1, μ_2, μ_3)		
$(\sigma_1, \sigma_2, \sigma_3)$	Theoretical VUS	(1, 2, 3)	(1, 1.5, 3)
	(0.5, 0.5, 0.5)	0.4124169 (0.945953519*) (0.95657799**)	0.3913009 (0.910320269*) (0.927020809**)
	(1.25,1.25, 1.25)	0.3029789 (0.761276894*) (0.793745033**)	0.2980189 (0.752906894*) (0.785716899**)
	(3, 3, 3)	0.2158898 (0.614314038*) (0.64149882**)	0.2152842 (0.613292088*) (0.640341574**)

* linear mapping ** quadratic mapping

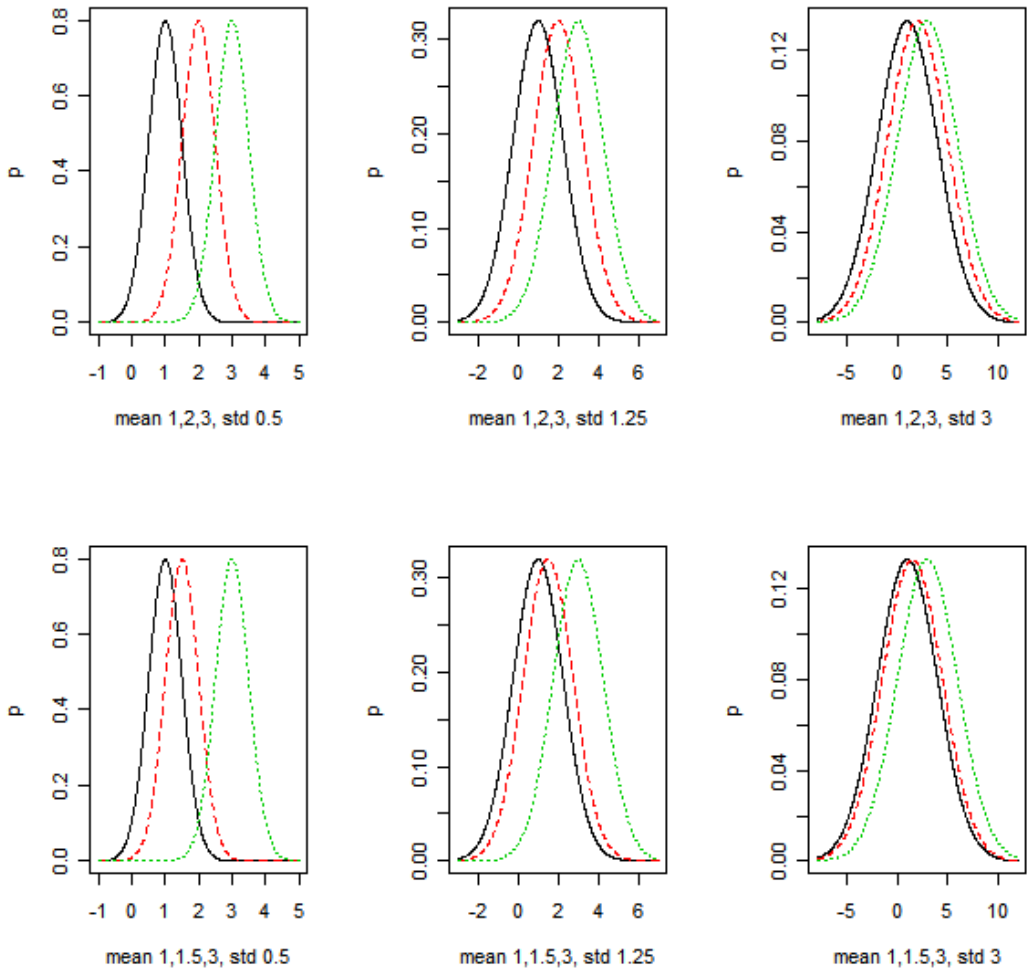

Figure 5.2 Population density for three normal distributions

Table 5.3 Theoretical VUS for gamma distribution

$\alpha_1 = \alpha_2 = \alpha_3 = 3$				
$(\beta_1, \beta_2, \beta_3)$	Beta	Theoretical VUS	Beta	Theoretical VUS
	(1, 4, 7)	0.370911 (0.875912313*) (0.897713108**)	(1, 5, 7)	0.3539129 (0.847228019*) (0.872658959**)
	(1, 2, 3)	0.3131569 (0.778452269*) (0.810015675**)	(1,2.5,3)	0.3026353 (0.760697069*) (0.793191019**)
	(1,1.25,1.5)	0.2128605 (0.609202094*) (0.635693736**)	(1,1.3,1.5)	0.2125777 (0.608724869*) (0.635149701**)

* linear mapping ** quadratic mapping

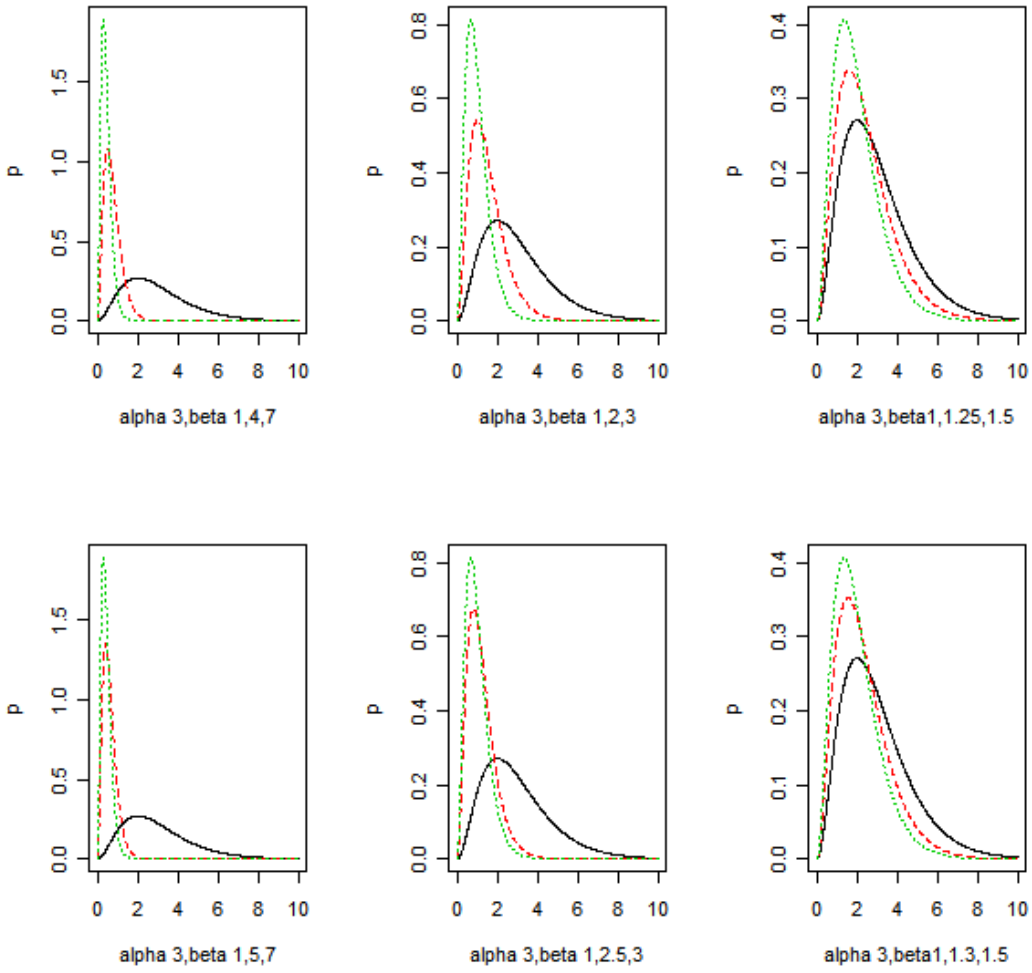

Figure 5.3 Population density for three gamma distributions

Table 5.4 Simulation for normal distribution

Parameter (μ_1,μ_2,μ_3) $(\sigma_1,\sigma_2,\sigma_3)$	Method	Sample size: (16, 8, 16)		Sample size: (40, 20, 40)	
		Coverage probability	Interval Length	Coverage probability	Interval Length
(1, 2, 3) (0.5, 0.5, 0.5)	U-Statistic	92.67%	0.0629859	93.33%	0.0412557
	Bootstrap	93.33%	0.0694444	93.00%	0.0358333
	EL	92.00%	0.0625617	92.67%	0.0314602
(1, 1.5, 3) (0.5, 0.5, 0.5)	U-Statistic	91.67%	0.0681728	94.33%	0.0545882
	Bootstrap	91.00%	0.0806944	92.67%	0.0555417
	EL	91.67%	0.0700987	92.33%	0.0471504
(1, 2, 3) (1.25, 1.25, 1.25)	U-Statistic	96.00%	0.1167010	95.00%	0.1042770
	Bootstrap	94.67%	0.1112196	93.33%	0.0954097
	EL	91.67%	0.0984592	93.00%	0.0892522
(1, 1.5, 3) (1.25, 1.25, 1.25)	U-Statistic	94.00%	0.1191885	95.33%	0.0963394
	Bootstrap	94.00%	0.1259592	93.00%	0.0801250
	EL	92.00%	0.0989351	92.67%	0.0774620
(1, 2, 3) (3, 3, 3)	U-Statistic	94.00%	0.1341069	95.33%	0.1108490
	Bootstrap	90.00%	0.1310764	94.33%	0.0936181
	EL	91.33%	0.1102676	92.67%	0.0869033
(1, 1.5, 3) (3, 3, 3)	U-Statistic	92.67%	0.1554159	94.67%	0.1274900
	Bootstrap	91.67%	0.1489800	91.33%	0.0978056
	EL	91.67%	0.1187315	92.67%	0.0853628

Table 5.5 Simulation for gamma distribution

Parameter $(\alpha_1, \alpha_2, \alpha_3)$ $(\beta_1, \beta_2, \beta_3)$	Method	Sample size: (16, 8, 16)		Sample size: (40, 20, 40)	
		Coverage probability	Interval Length	Coverage probability	Interval Length
(3, 3, 3) (1, 4, 7)	U-Statistic	91.00%	0.0715118	94.00%	0.0426619
	Bootstrap	93.33%	0.0876389	93.33%	0.0465625
	EL	90.67%	0.0706879	91.67%	0.0401577
(3, 3, 3) (1, 5, 7)	U-Statistic	91.67%	0.1089553	93.33%	0.0482550
	Bootstrap	92.00%	0.1158681	93.33%	0.0499583
	EL	91.33%	0.0867319	92.67%	0.0427180
(3, 3, 3) (1, 2, 3)	U-Statistic	96.00%	0.1254224	95.33%	0.0698427
	Bootstrap	94.00%	0.1379861	96.67%	0.0745069
	EL	92.00%	0.1138948	93.00%	0.0626753
(3, 3, 3) (1, 1.5, 3)	U-Statistic	95.00%	0.1013915	95.33%	0.0711492
	Bootstrap	93.00%	0.1156337	92.67%	0.0766806
	EL	91.67%	0.0959068	92.67%	0.0699184
(3, 3, 3) (1, 1.25, 1.5)	U-Statistic	94.33%	0.1334723	94.67%	0.0894067
	Bootstrap	92.67%	0.1402648	94.00%	0.0949444
	EL	91.33%	0.1175874	92.00%	0.0831169
(3, 3, 3) (1, 1.3, 1.5)	U-Statistic	95.33%	0.1530327	94.33%	0.0931584
	Bootstrap	93.00%	0.1425781	94.67%	0.0992847
	EL	91.33%	0.1289447	92.33%	0.0884236

CHAPTER 6

Introduction of a study on intellectual outcome in pediatric brain-tumor patients

6.1 Background:

Intracranial tumors are the second most common neoplasm occurring in children under 15 years of age with an estimated incidence of 1200 cases per year in the United States (Cohen & Duffner, 1984). While the absolute number of children with intracranial tumors is small, improved treatment outcomes (Finlay, Uteg, & Giese, 1987) have resulted in a need for increased attention to the quality of survival for these patients. The critical location of these neoplasms and the risk to cerebral integrity as a result of standard treatment that may include surgery, high-dose central nervous system (CNS) radiation, and/or neurotoxic chemotherapeutic agents put such children at greater risk for suboptimal behavioral, emotional, and cognitive outcomes as compared to children with other malignancies (Danoff Cowchock, Marquette, Mulgrew, & Kramer, 1982; Duffner, Cohen, & Thomas,1983; Hirsch, Renier, 1979; Le-Baron, Zeltzer, Scott, & Marlin, 1988).

Investigators are seeking a better understanding of the degree and type of short- and long-term neuropsychological impairment as well as important disease treatment, and patient risk factors related to this morbidity. Such information would be critical throughout treatment so that children and their families could be informed regarding anticipated outcomes and the development of effective rehabilitation programs could be facilitated. These data might also permit modification of treatment protocols to reduce subsequent morbidity without sacrificing

efficacy (Halberg et al., 1990). In addition, these children provide an opportunity to learn more about the long-term neuropsychological sequelae for a host of CNS trauma related to the neoplasms and their treatment.

The Stanford-Binet Intelligence Scale-Fourth Edition (SB-IV) [41] is a widely standardized measurement of intellectual functioning appropriate for individuals aged 2-23 years. Administration of the SB-IV yields estimates of overall intellectual functioning, verbal reasoning, quantitative reasoning, abstract visual reasoning, and short-term memory abilities. The indices of the SB-IV have demonstrated good reliability, internal consistency, and criterion-related validity across studies. In this dissertation paper, SB-IV Test is taken for each patient in several time points after diagnosis. Besides composite IQ, other four test scores (STB Verbal Reasoning SAS score, STB Abstract/Visual Reasoning SAS score, STB Quantitative Reasoning SAS score, STB Short-Term Memory SAS score) are also considered as response variables so that we can obtain more details about how the risk factors impact the performance of pediatric brain-tumor patients in tasks of intellectual functioning.

The children who have undergone treatment for brain tumors which have direct impact on crucial brain structures underlying behavior and may be more likely to exhibit cognitive difficulties than their peers. Although studies have found that survivors are at risk for a variety of physical, medical, cognitive, and/or psychosocial late effects, the particular risk factors having an impact on children's psychosocial and behavioral functioning are not fully understood. These late effects may be directly related to the type of treatment (surgery, chemotherapy, and/or radiation), characteristics of the disease (tumor size and type), and individual demographic factors, such as age and socioeconomic status.

For treatment factors, two types of measurements are taken into account and compared in this study. If the patients have taken radiation therapy and/or chemotherapy is considered as first type of treatment factor. Radiation therapy and chemotherapy are generally used as secondary or adjuvant treatments for tumors that cannot be managed using only surgery. However, radiation and chemotherapy may be used without surgery if the tumor is inoperable. Radiation therapy uses high-energy x-rays or other types of ionizing radiation to stop cancer cells from dividing. Because the developing brain of a child is very sensitive to radiation therapy, it is deliberately limited. Chemotherapy required for the more aggressive tumors uses chemicals (drugs) that have a toxic effect on tumor cells as they divide. Survival rates of children with certain types of brain tumors have been significantly improved by the treatment of radiation therapy and chemotherapy.

The Neurological Predictor Scale (NPS) [16] was used to look at multiple factors related to outcome. NPS was developed to address the needs of researchers in the field of pediatric neuro-oncology. It is a nonratio, ordinal scale that can be used to rate patients across 4 primary domains including tumor-related conditions, operative events, radiation treatment, and chemotherapy. Participants' ratings are summed across the 4 domains to yield a total score. The Neurological Predictor Scale can be calculated from a brief review of medical record data.

6.2 Source of data:

The data for this study comes from a longitudinal study conducted by Robin Morris of Georgia State University over 15 years ago. Tricia King in collaboration with Robin Morris and other researchers are evaluating the survivors of childhood brain tumors from the original longitudinal study when began at the time when children were diagnosed and treated. Drs. Tricia King and Robin Morris (Department of Psychology) and along with Dr. Yu-Sheng Hsu

(Mathematics and Statistics Department) are conducting studies to identify the predictors of longitudinal data such as the SB-IV test score. We analyzed change in five SB-IV test scores over time in these children.

Between 1985 and 1996, 153 patients participated in the longitudinal study and 95 out 153 patients are involved in this study whose information is complete and number of observations is at least two. The total number of observations among this 95 patients is 488 in which 478 STB Verbal Reasoning SAS scores, 476 STB Abstract/Visual Reasoning SAS scores, 478 STB Quantitative Reasoning SAS scores, 486 STB Short-Term Memory SAS scores and 486 STB Composite IQ scores are recorded. The patients' data also includes date of birth of the participant, gender, socioeconomic status, treatments (Radiation and/or Chemotherapy) the patient undertook, Neurological Predictor Scale the patient had, date of the diagnosis, and date of taking SB-IV test. The age of diagnosis of those 95 patients is ranged from 0.42 to 16.67 years old. The amount of time between diagnosis and SB-IV test in years is ranged from 0 to 15.92. The Neurological Predictor Scale is ranged from 0 to 10. The socioeconomic status is ranged from 1 to 5. The range of observation per patient is from 2 to 11.

The potentially predictive variables included in this study are gender, age at diagnosis, Socioeconomic Status Class (SES), chemotherapy, radiation, time since treatment and Neurological Predictor Scale (NPS). A family's socioeconomic status is based on family income, parental education level, and parental occupation. There are five levels for SES class in which class1 is the highest level and class 5 is the lowest level. Neurological Predictor Scale (NPS) is a nonratio, ordinal scale. It is a sum of patients' rated scores across 4 domains which are tumor-related conditions, operative events, radiation treatment, and chemotherapy [16].

Table 6.1 Descriptive table of treatments, gender, and SES classes

Variables	Patients					Observations				
	With		without			with		Without		
Radiation	56		39			282		206		
Chemotherapy	26		69			144		344		
	Patients					Observations				
Male	51					276				
Female	44					212				
Age<=7 years old	55					323				
Age>7 years old	40					165				
	Patients					Observations				
SES classes	1	2	3	4	5	1	2	3	4	5
	10	20	27	26	12	51	126	134	124	53

Table 6.2 Descriptive table of SB-IV scores, NPS, age at diagnosis and time

Variables	Mean	Standard Deviation	Range
STB Verbal Reasoning SAS scores	96.5188285	16.0424836	51 to 141
STB Abstract/Visual Reasoning SAS scores	94.8634454	18.7858919	38 to 191
STB Quantitative Reasoning SAS scores	93.0355649	16.4850733	56 to 136
STB Short-Term Memory SAS scores	92.9300412	18.6529687	50 to 153
STB Composite IQ scores	93.2489712	17.339834	50 to 140
Neurological Predictor Scale	5.82786885	2.12714005	0 to 10
Age at diagnosis (years)	6.18739754	3.60554068	0.42 to 16.67
Time(years between diagnosis and SB-IV test)	3.69686475	3.10603926	0 to 15.92

The data consisted of 488 records on 95 individuals. The frequency distribution of the number of time points is seen in Table 6.3. The five SB-IV scores on 95 children are displayed graphically in Fig. 6.1 - Fig. 6.5.

Table 6.3 Time point count distribution

Number of Time Points	Frequency	Percent	Cumulative Percent
2	15	15.79	15.79
3	12	12.63	28.42
4	18	18.95	47.37
More than 3	50	52.53	100.00

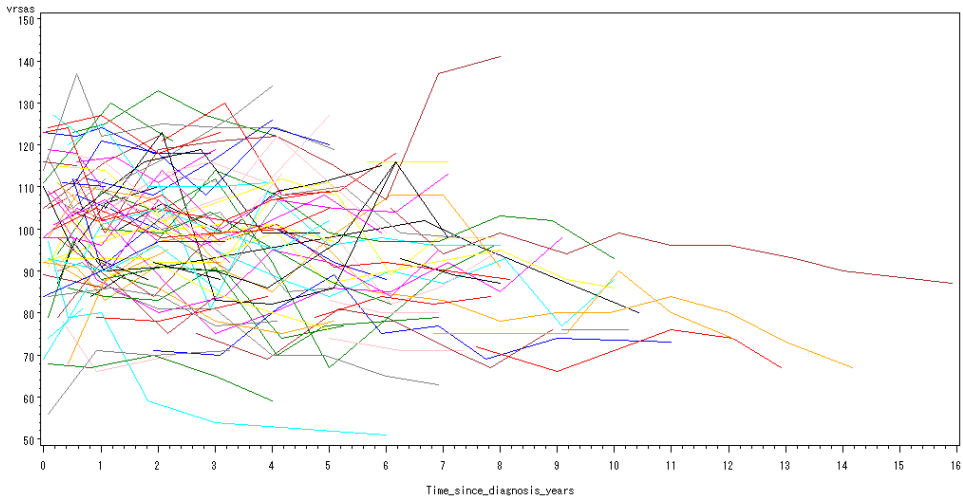

Figure 6.1 Individual Verbal Reasoning SAS score trajectories

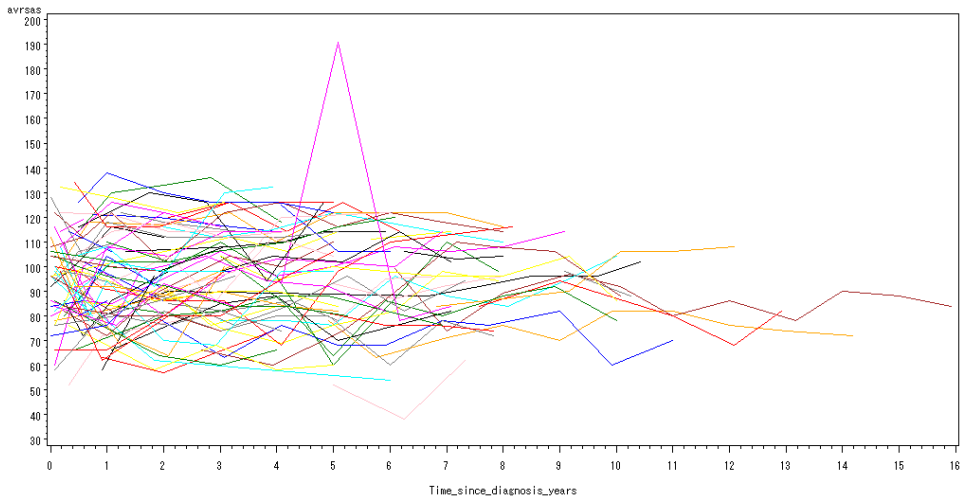

Figure 6.2 Individual Abstract/Visual Reasoning SAS score trajectories

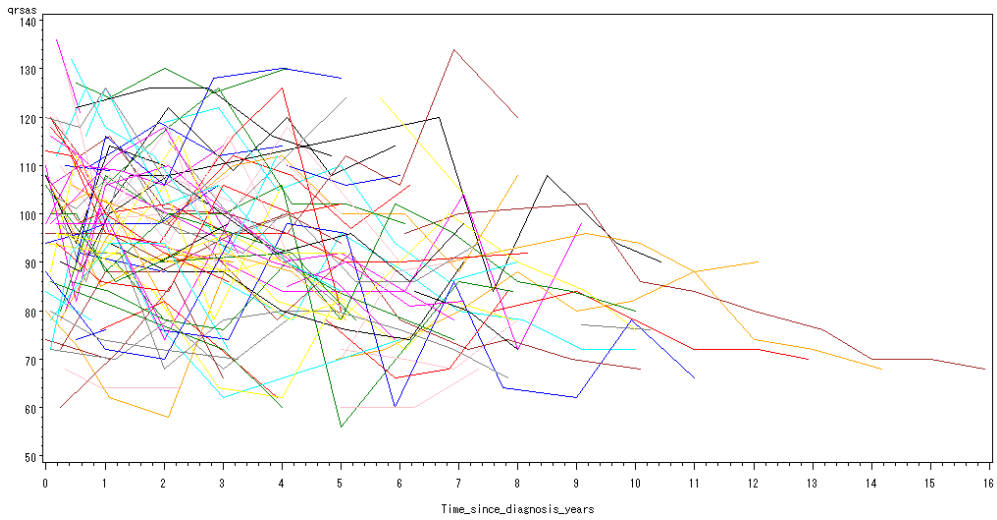

Figure 6.3 Individual Quantitative Reasoning SAS score trajectories

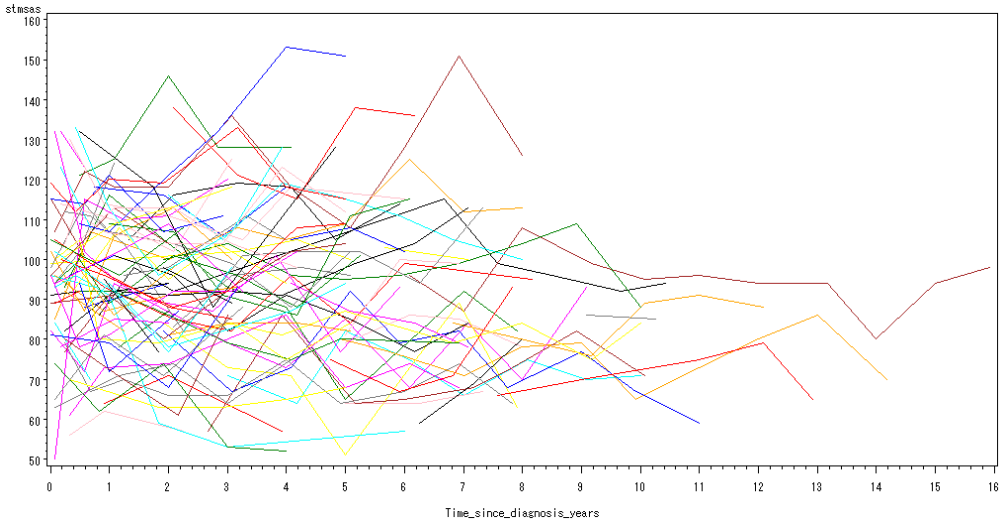

Figure 6.4 Individual Short-Term Memory SAS score trajectories

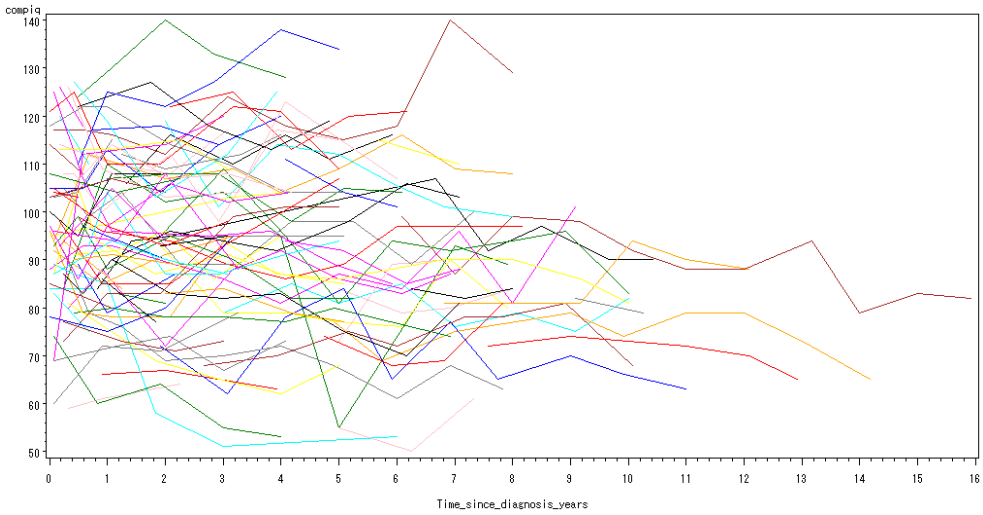

Figure 6.5 Individual composite IQ score trajectories

6.3 Method of analysis

In order to analyze change over time in psychological studies, there are numerous traditional methods that can be applied. These include the mixed model analysis of variance (ANOVA) and the multivariate approach to repeated measures, the analysis of covariance (ANCOVA) or residualized change analysis, and the analysis of covariance with reliability correction (ANCOVARC).

In this study, we will use hierarchical linear individual growth model and Gompertz based growth model to analyze the changes over time in SB-IV test score data. Much study shows that it is both possible and desirable to model change at the individual level [17]. The hierarchical linear individual growth model is a relative new statistical technique now widely used to examine the unique trajectories of individuals and groups in repeated measures data. And in this dissertation, we also propose a Gompertz based growth model. It is a nonlinear individual growth model, which turns out to be a better model than the hierarchical models for most of our response varibles using PRESS and VUS as criteria.

CHAPTER 7

Methodology and data analysis

In this chapter, the individual growth models are presented. Since repeated SB-IV test measurements taken on each child were obtained over time, hierarchical models were used for the analysis of change. To explain the change over time of five SB-IV test scores, two sets of explanatory variables were considered. One set includes Years between date of diagnosis and date of test (Time), Gender, Age group at diagnosis (Age group is define to be 1 if patient's age at diagnosis is greater than 7, otherwise, it is defined to be 0), SES classes, Treatments (radiation and chemotherapy), and potential interactions of these variables. The other model consists of Years between date of diagnosis and date of test (Time), Gender, Age group at diagnosis, SES classes, Neurological Predictor Scale (NPS), and potential interactions of these variables. The patients are assumed to be random and other variables are fixed effects in the model. In this chapter, both hierarchical linear model and Gompertz based nonlinear model are applied to the dataset and PRESS and VUS are selected as criteria to compare the performance of the models.

7.1 The hierarchical linear model (HLM)

Longitudinal studies sometimes known as repeated measures are encountered in a wide variety of disciplines. Longitudinal data is the union of cross-sectional and time series data. The balanced design in longitudinal data analysis assumes a complete data set with an equal number of measurements over time for each subject, while the unbalanced design has incomplete data

without equal time intervals or time points for each subject [18,19]. Literature shows that hierarchical linear model (HLM) can be employed in longitudinal data analysis.

When HLM is applied to longitudinal data analysis, the level 1 units are the repeated measures for each subject and the level 2 units consist of subjects. The repeated measures are conceived as nested within each subject. The level 1 model includes time or/and quadratic time as the predictor(s). The within-subject model is:

$$Y_{it} = \pi_{0i} + \pi_{1i} T_{it} + e_{it} \tag{7.1.1}$$

$$\text{or} \quad Y_{it} = \pi_{0i} + \pi_{1i} T_{it} + \pi_{2i} T_{it}^{\,2} + e_{it} \tag{7.1.2}$$

By convention, within person effects are indicated by the symbol π. Y_{it} represents the outcome for individual i measured at time t. T_{it} represents time from the base line assessment for person i. For model 1, the slope π_{1i} is the linear growth rate for the i^{th} person and the intercept, π_{0i}, represents the expected outcome of the person at baseline, also called initial status. For model 2, π_{1i} and π_{2i} are the coefficients of the first and second order of T_{it} respectively. The within-person residuals, e_{it}, are assumed $N(0, \sigma^2)$.

At level 2, the goal is to investigate variations in the estimates of the coefficients in level 1 model. The between-subjects models are:

$$\begin{aligned} \pi_{0i} &= \beta_{00} + \beta_{01} x_{1i} + ... + \beta_{0p-1} x_{p-1i} + u_{0i} \\ \pi_{1i} &= \beta_{10} + \beta_{11} x_{1i} + ... + \beta_{1p-1} x_{p-1i} + u_{1i} \\ \pi_{2i} &= \beta_{20} + \beta_{21} x_{1i} + ... + \beta_{2p-1} x_{p-1i} \end{aligned} \tag{7.1.3}$$

Accordingly, β_{00}, β_{10} and β_{20} represent the expected baseline and slope, respectively. The coefficients for the predictors indicate how much these expected values increase or decrease. The random effects at level 2 are assumed to be

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} - N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{00}^2 & \sigma_{01} \\ \sigma_{10} & \sigma_{11}^2 \end{pmatrix}\right] \tag{7.1.4}$$

Substitute Equation (7.1.3) into (7.1.1) or (7.1.2), we can reduce the 2-level model.

The hierarchical linear model extends the general linear model by allowing a more flexible specification of the covariance matrix of error. It allows for both correlation and heterogeneous variances.

The hierarchical linear model can be written as the form of mixed model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \varepsilon \tag{7.1.5}$$

Where **y** denotes the vector of observed values, **X** is the known matrix of explanatory variables, β is the unknown fixed-effect parameter vector, **Z** is the known design matrix of random effects, γ is the unknown random parameter and ε is the unobserved vector of independent and identically distributed Gaussian random errors.

A key assumption in the foregoing analysis is that γ and ε are normally distributed with

$$E\begin{bmatrix} \gamma \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } Var\begin{bmatrix} \gamma \\ \varepsilon \end{bmatrix} = \sigma^2 \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \tag{7.1.6}$$

The variance of y is, therefore, $\sigma^2 V = ZGZ' + R$. Here the structure of G is specified in (7.1.4) and (7.1.6) and $R = I_n$

After those assumptions, observed responses are treated as a realization of a multivariate Gaussian random vector, Y, with

$$Y \sim MVN(X\beta, \sigma^2 V) \tag{7.1.7}$$

Under the multivariate Gaussian assumption, the weighted least-square estimate $\hat{\beta}$ is the maximum likelihood estimator for β given V, i.e.,

$$\hat{\beta}(V) = (X'V^{-1}X)^{-1}X'V^{-1}y \tag{7.1.8}$$

Two likelihood-based methods: maximum likelihood (ML) and restricted/residual maximum likelihood (REML) can be implemented and the corresponding estimates of σ^2 are

$$\text{ML: } \hat{\sigma}^2(V) = RSS(V)/N \tag{7.1.9}$$

$$\text{REML: } \hat{\sigma}^2(V) = RSS(V)/(N-p) \tag{7.1.10}$$

where N is total sample size, p is the number of elements of β and

$$RSS(V) = (y - X\hat{\beta}(V))'V^{-1}(y - X\hat{\beta}(V)) \tag{7.1.11}$$

Substitute conditional estimates of β and σ^2 into log-likelihood functions, we can construct log-likelihood functions:

$$\text{ML: } l(V) = -\frac{1}{2}\log|V| - \frac{1}{2}RSS(V) - \frac{N}{2}\log(2\pi) \tag{7.1.12}$$

$$\text{REML: } l_R(V) = -\frac{1}{2}\log|V| - \frac{1}{2}RSS(V) - \frac{1}{2}\log|X'V^{-1}X| - \frac{N-p}{2}\log(2\pi) \tag{7.1.13}$$

Newton-Raphson algorithm is applied to maximize the log-likelihood function for estimate of V. In this dissertation, we use ML method to do estimation and inference.

7.2 Gompertz based hierarchical model

7.2.1 Gompertz nonlinear model

A Gompertz curve or Gompertz function, named after Benjamin Gompertz, is a sigmoid function. It is a commonly used mathematical model for a time series, where growth is slowest at the start and end of a time period.

$$y(t) = ae^{be^{ct}} \tag{7.2.1.1}$$

where *a* is the upper asymptote, *c* is the growth rate, and *b, c* are negative numbers.

In the sixties, Laird [20] for the first time successfully used the Gompertz curve to fit data of growth of tumors. In fact, tumors are cellular populations growing in a confined space where the availability of nutrients is limited. Denoting the tumor size as $X(t)$ it is useful to write the Gompertz Curve as follows:

$$y(t) = K\exp\left(\log(\frac{y(0)}{K})\exp(-\alpha t)\right) \tag{7.2.1.2}$$

where: $y(0)$ is the tumor size at the starting observation time; K is the carrying capacity, i.e. the maximum size that can be reached with the available nutrients. In fact it is: $\lim_{t\to+\infty} y(t) = K$ independently on $y(0) > 0$. Note that, in absence of therapies etc. usually it is $y(0) < K$, whereas, in presence of therapies, it may be $y(0) > K$; α is a constant related to the proliferative ability of the cells.

It is easy to verify that the dynamics of $y(t)$ is governed by the Gompertz differential equation:

$$y'(t) = \alpha\log(\frac{K}{y(t)})y(t) \tag{7.1.1.3}$$

Furthermore, $y'(t)$ can be expressed as: $y'(t) = F(y(t))y(t)$, where $F'(y) \le 0$ and $F(y)$ is the instantaneous proliferation rate of the cellular population, whose decreasing nature is due to the competition for the nutrients due to the increase of the cellular population, similarly to the logistic growth rate. However, there is a fundamental difference: in the logistic case the

proliferation rate for small cellular population is finite: $F(y) = \alpha(1 - (\frac{y}{K})^{v}) \Rightarrow F(0) = \alpha < +\infty$,

whereas in the Gompertz case the proliferation rate is unbounded:

$$\lim_{y \to 0+} F(y) = \lim_{y \to 0+} \alpha \log(\frac{K}{y}) = +\infty$$

As noticed by Steel [21] and by Wheldon [22] the proliferation rate of the cellular population is ultimately bounded by the cell division time. Thus, this might be an evidence that the Gompertz equation is not good to model the growth of small tumors. Moreover, more recently it has been noticed [23] that, including the interaction with immune system, Gompertz and other laws characterized by unbounded $F(0)$ would preclude the possibility of immune surveillance.

7.2.2 Gompertz based hierarchical model

In order to employ Gompertz function into longitudinal data analysis, we reparameterize the function and generalize it to hierarchical nonlinear model. Take log on both side of (7.2.1.1) and make simple reparameterization we can get log-Gompertz function:

$\log y(t) = \alpha_0 - \alpha_1 \exp(-\alpha_2 t)$. Similar to hierarchical linear model, we treat it as the 1 level model, i.e.,

$$\log Y_{it} = \alpha_{0i} - \alpha_{1i} \exp(-\alpha_{2i} t) + e_{it} \qquad (7.2.2.1)$$

At level 2, the goal is to investigate variations in the estimates of coefficients in level 1 model. The between-subjects models are:

$$\begin{aligned} \alpha_{0i} &= \beta_{00} + \beta_{01} x_{1i} + ... + \beta_{0p-1} x_{p-1i} + u_{0i} \\ \alpha_{1i} &= \beta_{10} + \beta_{11} x_{1i} + ... + \beta_{1p-1} x_{p-1i} \\ \alpha_{2i} &= \beta_{20} + \beta_{21} x_{1i} + ... + \beta_{2p-1} x_{p-1i} \end{aligned} \qquad (7.2.2.2)$$

$$\begin{aligned} & \alpha_{0i} = \beta_{00} + \beta_{01}x_{1i} + ... + \beta_{0p-1}x_{p-1i} + u_{0i} \\ \text{or} \quad & \alpha_{1i} = \beta_{10} + \beta_{11}x_{1i} + ... + \beta_{1p-1}x_{p-1i} + u_{1i} \\ & \alpha_{2i} = \beta_{20} + \beta_{21}x_{1i} + ... + \beta_{2p-1}x_{p-1i} \end{aligned} \tag{7.2.2.3}$$

The random effects at level 2 are assumed to be

$$u_{0i} \sim N(0, \sigma_{00}^2) \tag{7.2.2.4}$$

$$\text{or} \quad \begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} - N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{00}^2 & \sigma_{01} \\ \sigma_{10} & \sigma_{11}^2 \end{pmatrix}\right] \tag{7.2.2.5}$$

Substitute Equation (7.2.2.2) into (7.2.2.1) or (7.2.2.3) into (7.2.2.1), we can reduce the 2-level model.

The number of random effects can be selected arbitrary which depends on the criterion we select. In most of cases, one random effect can yield better weighted adjusted PRESS and VUS. In this dissertation paper, we select two-random-effect model only for short-term memory SAS score and for all other responses, we use one-random-effect model.

The Gompertz based hierarchical nonlinear model can be written as the form of nonlinear marginal mixed model:

$$\log \mathbf{y} = f(\beta) + \mathbf{Z}\gamma + \varepsilon \tag{7.2.2.6}$$

Where **y** denotes the vector of observed values, **f** is the known nonlinear vector function, β is the unknown fixed-effect parameter vector, **Z** is the known design matrix of random effects, γ is the unknown multi-normal random parameter with mean 0 and covariance matrix $\sigma^2 D$, and ε is the unobserved vector of independent and identically distributed Gaussian random errors with variance σ^2.

The key characteristic of this model is that the random effects enter the model in a linear fashion. This model is called marginal because the marginal expected value of the response variable, $\log y_i$ can be expressed in closed-form as a function of the population parameter β, namely, $E(\log y_i) = f_i(\beta)$. Consequently, even a straightforward Nonlinear Least Squares (NLS), which minimizes the sum of squares

$$\sum_{i=1}^{m} \| \log y_i - f_i(\beta) \|^2 , \tag{7.2.2.7}$$

produces a consistent estimator of β when $m \to \infty$ and $\{n_i\}$ are bounded.

Since γ and ε are normally distributed, the marginal model (7.2.2.6) can be written compactly as

$$\log y_i \sim N(f_i(\beta), \sigma^2(I + Z_i D Z_i')) \text{ , i=1,…,m} \tag{7.2.2.8}$$

with the log-likelihood function, up to a constant term $-\frac{N}{2}\ln(2\pi)$, with $N = \sum_{i=1}^{m} n_i$,

$$l(\theta) = -\frac{1}{2}\{N \ln \sigma^2 + \sum_{i=1}^{m}[\ln | I + Z_i D Z_i' | + \sigma^{-2}(\log y_i - f_i(\beta))'(I + Z_i D Z_i')^{-1}(\log y_i - f_i(\beta))]\} \tag{7.2.2.9}$$

where $\theta = (\beta, \sigma^2, vec(D))$ is the complete vector of parameters. The information matrix for the nonlinear model for θ has a block-diagonal form,

$$IF = \begin{bmatrix} \sum_{i=1}^{m} (\frac{\partial f_i}{\partial \beta})' V_i^{-1} (\frac{\partial f_i}{\partial \beta}) & 0 \\ 0 & H \end{bmatrix}, \tag{7.2.2.10}$$

where $H = \frac{1}{2}\sum_{i=1}^{m} \begin{bmatrix} n_i \sigma^{-4} & \sigma^{-2} vec'(R_i) D^{+\prime} \\ \sigma^{-2} D^{+} vec(R_i) & D^{+}(R_i \otimes R_i) D^{+\prime} \end{bmatrix}$, (7.2.2.11)

is the expected Hessian (information) matrix for variance parameters, and

$$V_i = I + Z_i D Z_i', \;\; V_i^{-1} = I - Z_i (D^{-1} + Z_i Z_i')^{-1} Z_i', \;\; D^+ = (D'D)^{-1} D' \tag{7.2.2.12}$$

$$R_i = Z_i (I + Z_i D Z_i')^{-1} Z_i' = ((Z_i' Z_i)^{-1} + D)^{-1}. \tag{7.2.2.13}$$

Note that the right-hand sides are valid if the matrices D and $Z_i' Z_i$ are invertible. The asymptotic covariance matrix for θ is the inverse of IF.

We may obtain a variance-profile log-likelihood,

$$l_p(\beta, D) = -0.5\{N \ln \sum (\log y_i - f_i(\beta)) V_i^{-1} (\log y_i - f_i(\beta)) + \sum \ln |V_i|\} \tag{7.2.2.14}$$

because when β and D are held fixed, the log-likelihood maximum is attained at $\sigma^2 = N^{-1} \sum (\log y_i - f_i(\beta)) V_i^{-1} (\log y_i - f_i(\beta))$. Since the information matrix, IF, has a block-diagonal form, we may maximize l over β and the variance parameters separately. Thus, the maximum likelihood estimate can be found using the following iterative algorithm [24]:

1. Set $D_0 = 0$ and apply nonlinear least squares, to find $\hat{\beta}_0$ and compute the residuals $\hat{e}_i = \log y_i - f_i(\hat{\beta}_0)$.

2. Find estimates for σ^2 and D by iterating until convergence,

$$\begin{bmatrix} \sigma^2 \\ vech(D) \end{bmatrix}_{s+1} = \begin{bmatrix} \sigma^2 \\ vech(D) \end{bmatrix}_s + \lambda_s H_s^{-1} \begin{bmatrix} \partial l / \partial \sigma^2 \\ vech(\partial l / \partial D) \end{bmatrix}, \tag{7.2.2.15}$$

where s is the iteration index and λ_s is a positive step length (typically $\lambda_s = 1$). Matrix H is defined by (7.2.2.11), and the derivatives of l with respect to the variance parameters are given by

$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{2} N \sigma^{-2} + \frac{1}{2} \sigma^{-4} \sum e_i' V_i^{-1} e_i \tag{7.2.2.16}$$

$$\frac{\partial l}{\partial D} = -\frac{1}{2} \sum [Z_i' V_i^{-1} Z_i - \sigma^{-2} Z_i' V_i^{-1} e_i e_i' V_i^{-1} Z_i] \tag{7.2.2.17}$$

for $e_i = \hat{e}_i$.

3. Find the weighted NLS solution, $\hat{\beta}_1$, to

$$\hat{\beta}_1 = \min_{\beta} \sum_{i=1}^{m} (\log y_i - f_i(\beta))'(I + Z_i D Z_i')^{-1}(\log y_i - f_i(\beta)) \qquad (7.2.2.18)$$

where $D = \hat{D}$ from step 2. Compute $\hat{e}_i = \log y_i - f_i(\hat{\beta}_1)$

4. Return to step 2 if the convergence criterion is not met.

To predict the response, since for two random effects model,

$$E(y_{it}) = \exp[\alpha_{0i} - \alpha_{1i}\exp(-\alpha_{2i}t) + \frac{1}{2}(\sigma_{00}^2 + \exp(-2\alpha_{2i}t)\sigma_{11}^2 - 2\exp(-\alpha_{2i}t)\sigma_{01} + \sigma^2)] \qquad (7.2.2.19)$$

and for one random effect model

$$E(y_{it}) = \exp[\alpha_{0i} - \alpha_{1i}\exp(-\alpha_{2i}t) + \frac{1}{2}(\sigma_{00}^2 + \sigma^2)] \qquad (7.2.2.20)$$

we define

$$\hat{y}_{it} = \exp[\hat{\alpha}_{0i} - \hat{\alpha}_{1i}\exp(-\hat{\alpha}_{2i}t) + \frac{1}{2}(\hat{\sigma}_{00}^2 + \exp(-2\hat{\alpha}_{2i}t)\hat{\sigma}_{11}^2 - 2\exp(-\hat{\alpha}_{2i}t)\hat{\sigma}_{01} + \hat{\sigma}^2)] \qquad (7.2.2.21)$$

for two random effect model and

$$\hat{y}_{it} = \exp[\hat{\alpha}_{0i} - \hat{\alpha}_{1i}\exp(-\hat{\alpha}_{2i}t) + \frac{1}{2}(\hat{\sigma}_{00}^2 + \hat{\sigma}^2)] \quad \text{for one random effect model.} \qquad (7.2.2.22)$$

7.3 PRESS residuals and weighted adjusted PRESS (WAPRESS)

The residual is defined to be the difference between the predicted value evaluated through the estimated model, and the observed value. Since the estimated model is derived from all observations including the one we want to predict, the residual tends to be smaller than it is

supposed to be. This is what we defined as the "shrinkage". One way to find more reasonable residual is to take away the observation we want to predict in the process of estimating the model. In the other words, using n-1 (n is the sample size) observations to estimate the model leaving one observation out. The PRESS residual is the difference between the estimated value on this observation and the true observed value. PRESS statistic is defined to be the sum of square PRESS residuals, which includes all n observations. It can be expressed as

$$\sum_{i=1}^{n} (\hat{Y}_{(-i)} - Y_i)^2 \qquad (7.3.1)$$

where Y_i is the *i*th observation and $\hat{Y}_{(-i)}$ is the estimated Y_i without using the *i*th observation, i = 1, 2, …, n.

Since the dataset in this study is longitudinal data which has repeated SB-IV test measurements obtained from each patient over time. It is more reasonable to adjust the PRESS residuals by taking away the observations of one specified patients we want to predict in the process of estimating the model. In the other words, using $N - n_i$ (N is the sample size and n_i is the number of observations for the ith patients) observations to estimate the model leaving n_i observations out. And the adjusted PRESS residuals are the difference between the estimated values on those n_i observations and the true observed values. The adjusted PRESS statistic is defined to be the sum of square adjusted PRESS residuals, which includes all N observations. It can be express as

$$\sum_{j=1}^{N} (\hat{Y}_{(-i[j])} - Y_j)^2 \qquad (7.3.2)$$

where Y_j is the *i*th observation and $\hat{Y}_{(-i[j])}$ is the estimated Y_j without using the n_i observations from the ith patient to which the jth observation belongs, j= 1, 2, …, N.

Since number of observations for each patient is different, it is reasonable to weight the adjusted PRESS residuals by the number of observations for each patient. We define weighted adjusted PRESS (WAPRESS) as

$$WAPRESS = \sum_{i=1}^{m} \frac{1}{n_i} \sum_{j=1}^{n_i} (\hat{Y}_{(-i)j} - Y_{ij})^2 \qquad (7.3.3)$$

where Y_{ij} is the jth observation of ith patient, $j = 1,...n_i$ and $i = 1,...m$, and $\hat{Y}_{(-i)j}$ is the estimated Y_{ij} without using the n_i observations from the ith patient.

7.4 VUS

According to hierarchical linear model or Gompertz model, we can predict the SB-IV test scores for different time points of patients. In some case, the scores are grouped to different levels, such as “low”, “median”, “high”. Then VUS proposed in the first part of dissertation can be applied as criterion to evaluate the goodness of the models. Different from PRESS, the VUS of one model is higher if the predictions of observations are more correctly classified instead of closer to the true scores. The details of new proposed VUS criterion can be found in the first five chapters of this dissertation.

7.5 Statistical analysis

SAS PROC MIXED and PROC NLMIXED provides a very flexible ways to model many types of repeated measures data. In this dissertation, PROC MIXED is used to build hierarchical linear models on two sets of predictors and PROC NLMIXED is applied to construct Gompertz based hierarchical nonlinear models on two sets of predictors. Since the analysis procedure of the five response variables is similar, in this part, we only show the analysis of SB-IV composite IQ

score. The details of the other four responses are attached in the appendix. In this study, backward elimination is applied to do model selection.

7.5.1 Hierarchical linear model

7.5.1.1 Hierarchical linear model for treatments and other variables

The variables interested are Years between date of treatment and date of exam (Time), Gender, Age group, SES classes, Treatments (radiation and chemotherapy). The patients are divided into two age groups according to the age of diagnosis at cutting point seven-year-old. Based on Chemotherapy (0=No, 1=Yes), Radiation (0=No, 1=Yes), Age group (0: age at diagnosis is less than 7 years old, 1: otherwise), Gender (1=Female, 0=Male), and SES class (1, 2, 3, 4, 5), 80 categories are defined.

Two models are considered in this section: Model1: Linear model without interaction among risk factors; Model 2: Quadratic model without interaction. In this set of variables, for either linear model or quadratic model, no interaction is significant among risk factors.

Table 7.1 shows the solution for fixed effects of Model 1. Table 7.2 presents the random effects which is the variances of the intercept and linear slope.

Table 7.1 Fixed effects for the hierarchical linear model without interactions for treatments and other variables

Effect	Estimate	Standard Error	DF	Pr > |t|
Intercept	113.30	3.9511	92	<.0001
SES	-7.0601	1.1457	295	<.0001
age_g	8.3044	2.7355	295	0.0026
rad*time	-2.1324	0.4357	295	<.0001
SES*time	0.3167	0.1145	295	0.0061

Table 7.2 Random effects for the hierarchical linear model without interactions for treatments and other variables

Cov Parm	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	154.08	28.2940	5.45	<.0001
UN(2,1)	-0.8865	3.5055	-0.25	0.8004
UN(2,2)	1.5312	0.7334	2.09	0.0184

The final expression for Model 1 is

$$Compiq_{it} = (113.30 + 8.3044\,Age\_group_i - 7.0601 SES_i + \mu_{0i})$$

$$+ (-2.1324 Rad_{it} + 0.3167 SES_i + \mu_{1i}) Time_{it} + \varepsilon_{it} \qquad (7.5.1)$$

Table 7.3 shows the solution for fixed effects of Model 2. Table 7.4 presents the random effects which is the variances of the intercept and linear slope.

Table 7.3 Fixed effects for the hierarchical quadratic model without interactions for treatments and other variables

Effect	Estimate	Standard Error	DF	Pr > |t|
Intercept	114.74	3.7944	93	<.0001
SES	-6.4198	1.1372	293	<.0001
time2	0.07803	0.03204	293	0.0155
rad*time	-2.0928	0.3940	293	<.0001
time*age_g	3.1841	0.9433	293	0.0008
time2*age_g	-0.3518	0.1452	293	0.0160

Table 7.4 Random effects for the hierarchical quadratic model without interactions for treatments and other variables

Cov Parm	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	153.30	28.1212	5.45	<.0001
UN(2,1)	-0.8923	3.5294	-0.25	0.8004
UN(2,2)	1.5581	0.7340	2.12	0.0169

The final expression for Model 2 is

$$Compiq_{it} = (114.74 - 6.4198SES_i + \mu_{0i}) + (3.1841Age\_group_i - 2.0928Radiation_{it} + \mu_{1i})Time_{it} + (0.07803 - 0.3518Age\_group_i)Time_{it}^2 + \varepsilon_{it} \quad (7.5.2)$$

7.5.1.2 Hierarchical linear model for neurological predictor scale and other variables

The variables interested are Years between date of treatment and date of exam (Time), Gender, Age group, SES classes, Neurological Predictor Scale (NPS). The patients are divided into two age groups according to the age of diagnosis at cutting point seven-year-old.

Four models are considered in this section: Model3: Linear model without interaction among risk factors; Model 4: Linear model with interaction; Model5: Quadratic model without interaction among risk factors; Model 6: Quadratic model with interaction;

Table 7.5 shows the solution for fixed effects of Model 3. Table 7.6 presents the random effects which is the variances of the intercept and linear slope.

Table 7.5 Fixed effects for the hierarchical linear model without interactions for NPS and other variables

Effect	Estimate	Standard Error	DF	Pr > |t|
Intercept	117.43	4.5235	92	<.0001
SES	-6.2047	1.1087	294	<.0001
age_g	8.0850	2.6447	294	0.0024
NPS	-1.3286	0.5134	294	0.0101
time	2.0333	0.7307	94	0.0065
NPS*time	-0.3520	0.1121	294	0.0019

Table 7.6 Random effects for the hierarchical linear model without interactions for NPS and other variables

Cov Parm	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	147.84	27.3825	5.40	<.0001
UN(2,1)	-2.9432	3.7870	-0.78	0.4371
UN(2,2)	2.1696	0.8721	2.49	0.0064

The find expression for Model 3 is

$$Compiq_{it} = (117.43 - 6.2047 SES_i + 8.0850 Age\_group_i - 1.3286 NPS_{it} + \mu_{0i}) + (2.0333 - 0.3520 NPS_{it} + \mu_{1i}) Time_{it} + \varepsilon_{it} \quad (7.5.3)$$

Table 7.7 shows the solution for fixed effects of Model 4. Table 7.8 presents the random effects which is the variances of the intercept and linear slope.

The final expression of Model 4 is

$$Compiq_{it} = (108.57 + 9.0307 Age\_group_i - 10.4269 SES_i + 14.8456 Sex_i + 0.7480 NPS_{it} SES_i - 2.5682 NPS_{it} Sex_i + \mu_{0i}) + (2.0887 - 0.3600 NPS_{it} + \mu_{1i}) Time_{it} + \varepsilon_{it} \quad (7.5.4)$$

Table 7.7 Fixed effects for the hierarchical linear model with interactions for NPS and other variables

Effect	Estimate	Standard Error	DF	Pr > |t|
Intercept	108.57	5.1260	91	<.0001
SES	-10.4269	2.2729	293	<.0001
age_g	9.0307	2.6034	293	0.0006
Sex	14.8456	5.1080	293	0.0039
time	2.0887	0.7273	94	0.0050
SES*NPS	0.7480	0.3562	293	0.0366
Sex*NPS	-2.5682	0.7919	293	0.0013
time*NPS	-0.3600	0.1115	293	0.0014

Table 7.8 Random effects for the hierarchical linear model with interactions for NPS and other variables

Cov Parm	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	140.43	26.2541	5.35	<.0001
UN(2,1)	-3.1681	3.6042	-0.88	0.3794
UN(2,2)	2.1660	0.8608	2.52	0.0059

Table 7.9 shows the solution for fixed effects of Model 5. Table 7.10 presents the random effects which is the variances of the intercept and linear slope.

Table 7.9 Fixed effects for the hierarchical quadratic model without interactions for NPS and other variables

Effect	Estimate	Standard Error	DF	Pr > |t|
Intercept	117.40	4.4872	92	<.0001
SES	-6.3331	1.1068	292	<.0001
NPS	-1.1470	0.5440	292	0.0359
age_g	7.7608	2.6463	292	0.0036
time2	0.2486	0.1054	292	0.0190
Sex*time	1.2789	0.4249	292	0.0028
NPS*time	-0.3735	0.1117	292	0.0009
time2*Sex	-0.1485	0.06366	292	0.0204

Table 7.10 Random effects for the hierarchical quadratic model without interactions for NPSand other variables

Cov Parm	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	145.59	27.1064	5.37	<.0001
UN(2,1)	-2.4145	3.8207	-0.63	0.5274
UN(2,2)	2.1739	0.8807	2.47	0.0068

The final expression of Model 5 is

$$Compiq_{it} = (117.40 - 6.3331 SES_i - 1.1470 NPS_{it} + 7.7608 Age\_group_i + \mu_{0i}) + (-0.3735 NPS_{it} + 1.2789 Sex_i + \mu_{1i}) Time_{it} + (0.2486 - 0.1485 Sex_i) Time_{it}^2 + \varepsilon_{it} \quad (7.5.5)$$

Table 7.11 shows the solution for fixed effects of Model 6. Table 7.12 presents the random effects which is the variances of the intercept and linear slope.

Table 7.11 Fixed effects for the hierarchical quadratic model with interactions for NPS and other variables

Effect	Estimate	Standard Error	DF	Pr > |t|
Intercept	112.24	5.1395	91	<.0001
SES	-11.1305	2.2850	291	<.0001
age_g	8.9409	2.5964	291	0.0007
Sex	13.5522	5.0925	291	0.0082
time2	0.2507	0.1044	291	0.0169
SES*NPS	0.8660	0.3595	291	0.0166
Sex*NPS	-2.7321	0.7948	291	0.0007
NPS*time	-0.4151	0.1109	291	0.0002
Sex*time	1.4826	0.4348	291	0.0007
Sex*time2	-0.1483	0.06328	291	0.0198

Table 7.12 Random effects for the hierarchical quadratic model with interactions for NPS and other variables

Cov Parm	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	136.75	25.7335	5.31	<.0001
UN(2,1)	-2.5404	3.6020	-0.71	0.4806
UN(2,2)	2.1330	0.8533	2.50	0.0062

The final expression of Model 6 is

$$\begin{aligned} Compiq_{it} &= (112.24 - 11.1305SES_i + 8.9409Age\_group_i + 13.5522Sex_i \\ &+ 0.8660SES_iNPS_{it} - 2.7321Sex_iNPS_{it} + \mu_{0i}) \\ &+ (1.4826Sex_i - 0.4151NPS_{it} + \mu_{1i})Time_{it} \\ &+ (0.2507 - 0.1483Sex_i)Time_{it}^2 + \varepsilon_{it} \end{aligned} \quad (7.5.6)$$

7.5.2 Gompertz based hierarchical nonlinear model

7.5.2.1 Gompertz based hierarchical nonlinear model for treatments and other variables

The variables interested are Years between date of treatment and date of exam (Time), Gender, Age group, SES classes, Radiation and Chemotherapy. The patients are divided into two age groups according to the age of diagnosis at cutting point seven-year-old.

One model is considered in this section: Model7: Gompertz based hierarchical nonlinear model without interaction among risk factors. For this model, no interaction is significant among risk factors.

Table 7.13 shows the solution for fixed effects and variance of random effects of Model 7.

Table 7.13 Fixed effects and variance of random effect for the Gompertz based nonlinear hierarchical model without interactions for treatments and other variables

Parameter	Estimate	Standard Error	DF	Pr > |t|
beta00	4.7174	0.04348	94	<.0001
beta01	0.1320	0.03270	94	0.0001
beta02	-0.06447	0.01244	94	<.0001
beta03	-0.1192	0.02879	94	<.0001
beta11	0.1071	0.03021	94	0.0006
beta12	-0.1840	0.02512	94	<.0001
beta21	0.3861	0.1673	94	0.0232
sigmae	0.006364	0.000456	94	<.0001
sigmau1	0.01862	0.002943	94	<.0001

The final expression for model 7 is

$$\begin{aligned}\log(Compiq_{it}) &= (4.7174 + 0.1320 Age\_group_i \\ &- 0.06447 SES_i - 0.1192 Radiation_{it} + \mu_{0i}) \\ &- (0.1071 Age\_group_i - 0.1840 Radiation_{it}) e^{-(0.3861 Sex_i) Time_{it}} + \varepsilon_{it}\end{aligned} \tag{7.5.7}$$

7.5.2.2 Gompertz based hierarchical nonlinear model for neurological predictor scale and other variables

The variables interested are Years between date of treatment and date of exam (Time), Gender, Age group, SES classes, Neurological Predictor Scale (NPS). The patients are divided into two age groups according to the age of diagnosis at cutting point seven-year-old.

2 models are considered in this section: Model8: Gompertz hierarchical nonlinear model without interaction among risk factors; Model 9: Gompertz hierarchical nonlinear model with interaction.

Table 7.14 shows the solution for fixed effects and variance of random effects of Model 8.

Table 7.14 Fixed effects and variance of random effect for the Gompertz based nonlinear hierarchical model without interactions for NPS and other variables

Parameter	Estimate	Standard Error	DF	Pr > |t|
beta00	4.9967	0.05623	94	<.0001
beta01	0.07350	0.02751	94	0.0089
beta02	-0.08743	0.01258	94	<.0001
beta03	-0.04390	0.006028	94	<.0001
beta10	0.2742	0.04944	94	<.0001
beta11	-0.02577	0.008297	94	0.0025
beta12	-0.04275	0.006383	94	<.0001
beta21	-0.2538	0.03815	94	<.0001
beta22	0.2455	0.07762	94	0.0021
beta23	0.1381	0.02457	94	<.0001
sigmae	0.006099	0.000437	94	<.0001
sigmau1	0.01577	0.002525	94	<.0001

The final expression for model 8 is

$$\log(Compiq_{it}) = (4.9967 + 0.0735 Age\_group_i - 0.08743 SES_i - 0.0439 NPS_{it} + \mu_{0i}) - (0.2742 - 0.02577 SES_i - 0.04275 NPS_{it}) e^{-(-0.2538 SES_i + 0.2455 Sex_i + 0.1381 NPS_{it}) Time_{it}} + \varepsilon_{it} \quad (7.5.8)$$

Table 7.15 shows the solution for fixed effects and variance of random effects of Model 9.

The final expression for model 9 is

$$\log(Compiq_{it}) = (5.0313 - 0.09848 SES_i - 0.0437 NPS_{it} + 0.0228 Age\_group_i SES_i + \mu_{0i}) - (0.2747 - 0.02589 SES_i - 0.04275 NPS_{it}) e^{-(-0.2538 SES_i + 0.2479 Sex_i + 0.1376 NPS_{it}) Time_{it}} + \varepsilon_{it} \quad (7.5.9)$$

Table 7.15 Fixed effects and variance of random effect for the Gompertz based nonlinear hierarchical model with interactions for NPS and other variables

Parameter	Estimate	Standard Error	DF	Pr > |t|
beta00	5.0313	0.05424	94	<.0001
beta01	-0.09848	0.01306	94	<.0001
beta02	-0.04370	0.006028	94	<.0001
beta03	0.02280	0.008292	94	0.0072
beta10	0.2747	0.04941	94	<.0001
beta11	-0.02589	0.008285	94	0.0024
beta12	-0.04275	0.006382	94	<.0001
beta21	-0.2538	0.03810	94	<.0001
beta22	0.2479	0.07798	94	0.0020
beta23	0.1376	0.02443	94	<.0001
sigmae	0.006100	0.000437	94	<.0001
sigmau1	0.01569	0.002515	94	<.0001

CHAPTER 8

Model comparison and interpretation

In this chapter, we use WAPRESS and VUS as criterions to compare the nine models about composite IQ presented in chapter 7. Also the relationship between the risk factors and composite IQ for the nine models are discussed respectively. The comparison and interpretation for the other four responses can be found in the appendix.

8.1 Model comparison using WAPRESS

Table 8.1 presents the weighted adjusted PRESS for nine models. No interaction is significant for radiation & chemotherapy models.

From Table 8.1 we found that for all cases, Gompertz based hierarchical nonlinear models yield better WAPRESS results than hierarchical linear/quadratic models. And furthermore, Gompertz improved more for NPS models. It shows the merit of Gompertz structure.

For the two sets of predictors, Neurological Predictor Scale models always perform better than radiation & chemotherapy models. It is because Neurological Predictor Scale includes more information than simple radiation and chemotherapy model. Then it can be an evidence that NPS is good criterion for measure the patients' treatment condition and medical complications.

And we can see in hierarchical linear models, adding quadratic terms does not seem to have significantly better results. But for NPS models, models with interaction yields better WAPRESS than the models without interaction.

Table 8.1 WAPRESS for composite IQ models

Without interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	19591.82	19335	19055.68
Neurological Predictor Scale	18857.45	19091.45	17149.66

With interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	NA	NA	NA
Neurological Predictor Scale	18313.01	18506.5	17136.92

8.2 Model comparison using VUS

Weighted Adjusted PRESS measures how close the predicted response to the observed one. Sometimes we are more interested in classifying the response variable into several ordered levels, such as our grade system in schools of this country. VUS can be used to measure and compare the models constructed in previous section. Since the IQ score in this study is for the pediatric patients whose IQ will be lower than the common population, the regular classification should not be used to group the IQ score. In this study, we assume the IQ is normal and group the IQ into three and four levels such that for each level, the population is approximately even. Then the cutoff point for 3-level case is: Low: <86, Median: 86-100, High: >100. The cutoff point for 4-level case is Low: <82, Median low: 82-93, Median high: 94-104, High: >104.

We use the predictive value in the nine models stated in chapter 2 and the output of 3-level VUS and linear/quadratic transformed VUS are presented in table 8.2 and 4-level VUS and linear/cubic transformed VUS are presented in table 8.3. These two tables show the performance of the models corresponding to their power of classification. It is not hard to notice the advantage

of Gompertz model structure. It is also interesting to note that the prediction is better for 3-level responses than that of 4-level responses. This might indicate that it is harder to predict more precise IQ scores.

Table 8.2 3-level VUS and transformed VUS for composite IQ models

Without interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	0.3589856 (0.749521) * (0.78737) **	0.3526003 (0.741662) (0.779826)	0.3633188 (0.754854) (0.792452)
Neurological Predictor Scale	0.3787719 (0.773873) (0.810333)	0.377479 (0.772282) (0.808851)	0.402626 (0.803232) (0.837233)

With interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	NA	NA	NA
Neurological Predictor Scale	0.379855 (0.775206) (0.811573)	0.3765946 (0.771193) (0.807836)	0.4003582 (0.800441) (0.834711)

(* linear transformed VUS **quadratic transformed VUS)

Table 8.3 4-level VUS and transformed VUS for composite IQ models

Without interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	0.0881244 (0.654734) * (0.734454) **	0.0850012 (0.646739) (0.7257)	0.0938991 (0.669517) (0.750109)
Neurological Predictor Scale	0.0955790 (0.673818) (0.754542)	0.094217 (0.670331) (0.750952)	0.1060152 (0.700535) (0.78098)

With interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	NA	NA	NA
Neurological Predictor Scale	0.0949701 (0.672259) (0.752941)	0.0944536 (0.670937) (0.751578)	0.1049884 (0.697906) (0.778458)

(* linear transformed VUS **cubic transformed VUS)

8.3 Model interpretation

In this section, the effect of each risk factor to the composite IQ score will be discussed from the results of the nine models stated in chapter 2. Chemotherapy is not significant in all nine models, and then here five risk factors (Age group, SES, Sex, Radiation, NPS) are analyzed separately.

8.3.1 Effect of age group

Table 8.4 shows the effect of older age at diagnosis (more than 7 years old at diagnosis). We can find individuals whose age is 7 years old or older at diagnosis will have higher composite IQ than those whose age is younger than 7 years old in eight models. Treatment hierarchical quadratic model shows that the benefit of older age at diagnosis only lasts 9 years after diagnosis. Gompertz based hierarchical nonlinear models also provide more information: the benefit of older age at diagnosis is larger in female than male (see figure 8.1); the lower socioeconomic status class, the more benefit of older age at diagnosis (see figure 8.2).

In figure 8.1, composite IQ of patients whose age is less than 7 years old and more than 7 years old at diagnosis are compared in both male and female patients who have SES=3 with radiation. In figure 8.2, composite IQ of patients whose age is less than 7 years old and more than 7 years old at diagnosis are compared in both low SES (SES=5) and high SES (SES=1) male patient with NPS=8.

8.3.2 Effect of socioeconomic status class

Table 8.5 shows the effect of socioeconomic status class to composite IQ score. Most of the models show that low SES is associated with lower composite IQ. And from the treatment

Table 8.4 Effect of age group to composite IQ score

Model	$\frac{\partial IQ}{\partial Age\_group}$ or $\frac{\partial \log IQ}{\partial Age\_group}$
Model 1 (treatment linear w/o inter)	8.3043
Model 2 (treatment quadratic w/o inter)	$3.1841t - 0.3518t^2$
Model 3 (NPS linear w/o inter)	8.085
Model 4 (NPS linear w/ inter)	9.0307
Model 5 (NPS quadratic w/o inter)	7.7608
Model 6 (NPS quadratic w/ inter)	8.9409
Model 7 (treatment Gompertz w/o inter)	$0.132 - 0.1071\exp(-0.3861 * Sex * t)$
Model 8 (NPS Gompertz w/o inter)	0.0735
Model 9 (NPS Gompertz w/ inter)	$0.0228SES$

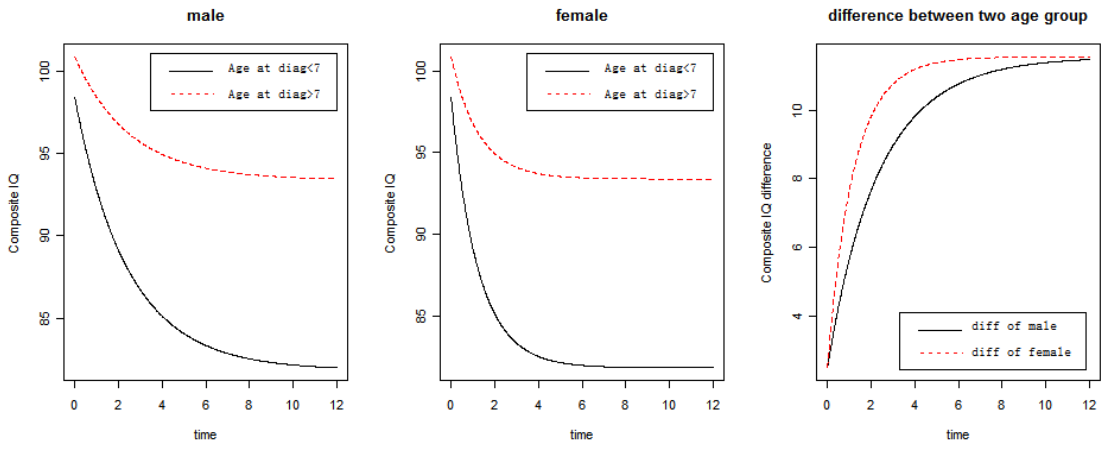

Figure 8.1 Composite IQ comparison between age groups in different sex in model 7

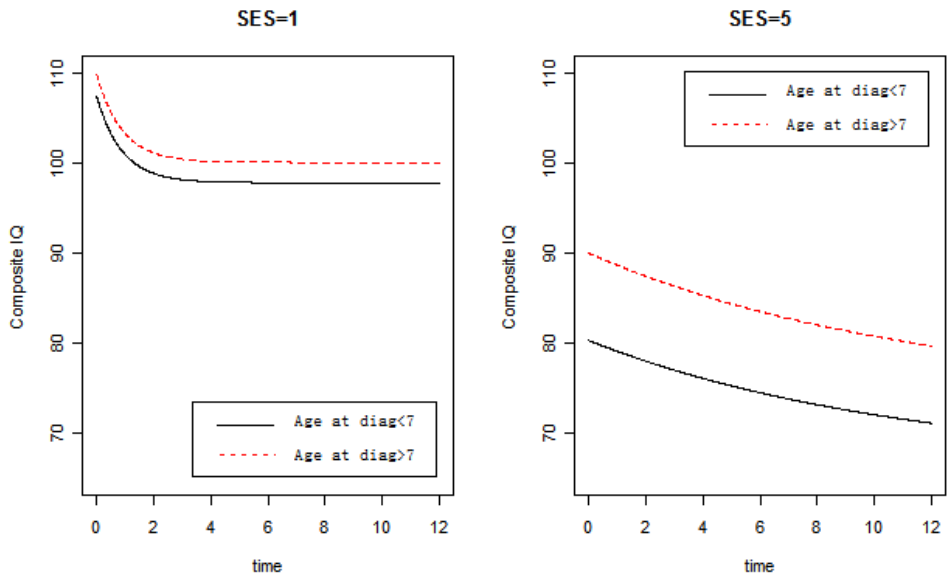

Figure 8.2 Composite IQ comparison between age groups in low and high SES in model 9

hierarchical linear model we can find as time passing, the harm will go smaller. And NPS hierarchical linear/quadratic models show the higher NPS score, the lower harm low SES will give (see figure 8.3). The interpretation of SES effect in NPS Gompertz based hierarchical nonlinear models is not clear due to the complex structure of the models.

In figure 8.3, composite IQ of patients who have low SES (SES=5) and high SES (SES=1) are compared in both low NPS (NPS=0) and high NPS (NPS=10) male patient whose age is more than 7 years old at diagnosis.

8.3.3 Effect of sex

Table 8.6 shows the effect of sex to composite IQ score. Sex is significant in six out of nine models. The hierarchical linear/quadratic models show that when NPS is low and time after

Table 8.5 Effect of socioeconomic status class to composite IQ score

Model	$\frac{\partial IQ}{\partial SES}$ or $\frac{\partial \log IQ}{\partial SES}$
Model 1 (treatment linear w/o inter)	$-7.0601+0.3167t$
Model 2 (treatment quadratic w/o inter)	-6.4198
Model 3 (NPS linear w/o inter)	-6.2047
Model 4 (NPS linear w/ inter)	$-10.4269+0.748NPS$
Model 5 (NPS quadratic w/o inter)	-6.3331
Model 6 (NPS quadratic w/ inter)	$-11.1305+0.866NPS$
Model 7 (treatment Gompertz w/o inter)	-0.06447
Model 8 (NPS Gompertz w/o inter)	$-0.088+(0.026+0.007t+0.0065SES*t+0.01NPS*t)$ *exp{(0.2538SES - 0.2455Sex - 0.1381NPS)t}
Model 9 (NPS Gompertz w/ inter)	$-0.1+0.02Age\_group+(0.026+0.007t+0.0066SES*t+0.01NPS*t)$ *exp{(0.2538SES - 0.2479Sex - 0.1376NPS)t}

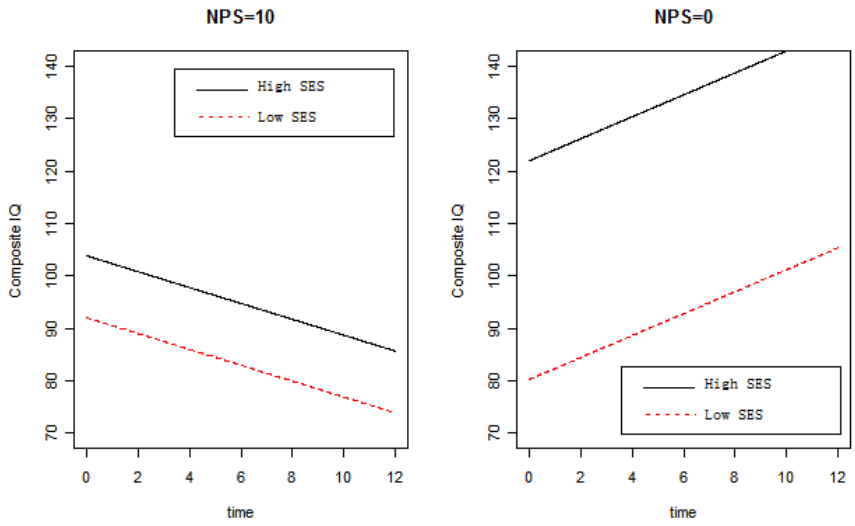

Figure 8.3 Composite IQ comparison between low and high SES in low and high NPS in model 4

diagnosis is long, male have higher composite IQ than female. Otherwise, when NPS is high and time after diagnosis is short, female have higher composite IQ. And treatment Gompertz based hierarchical nonlinear model shows that for those whose age is 7 years older at diagnosis and without radiation, female have higher composite IQ than male. For those whose age is 7 years younger at diagnosis and without radiation, female and male make no difference in composite IQ. For those who have radiation, female have lower composite IQ than male (see figure 8.4). NPS Gompertz based hierarchical nonlinear models show that when SES is high and NPS is low, male have higher composite IQ than female. Otherwise, when SES is low and NPS is high, female have higher composite IQ. If both SES and NPS are high or low, it depends. ($0.275 - 0.026SES - 0.043NPS > 0$, then female have higher composite IQ. Otherwise, male have higher composite IQ). (see figure 8.5)

In figure 8.4, composite IQ of male and female patients are compared in four combinations of patient whose age is less and more than7 years old at diagnosis with and without radiation in SES=3 .

In figure 8.5, composite IQ of male and female patients are compared in four combinations of low and high SES (SES=5 and 1) and low and high NPS (NPS=0 and 10) whose age is more than7 years old at diagnosis.

8.3.4 Effect of radiation

Table 8.7 shows the effect of radiation to composite IQ score. The hierarchical linear/quadratic models show that radiation is harmful for the patients' composite IQ and as time passing, the harm will increase. And treatment Gompertz based hierarchical nonlinear model shows that for female, 0.55 years after diagnosis, and for male, 1.1 years after diagnosis patients

Table 8.6 Effect of sex to composite IQ score

Model	$\frac{\partial IQ}{\partial Sex}$ or $\frac{\partial \log IQ}{\partial Sex}$
Model 1 (treatment linear w/o inter)	N/A
Model 2 (treatment quadratic w/o inter)	N/A
Model 3 (NPS linear w/o inter)	N/A
Model 4 (NPS linear w/ inter)	$14.8456 - 2.5682NPS$
Model 5 (NPS quadratic w/o inter)	$1.2789 - 0.1485t^2$
Model 6 (NPS quadratic w/ inter)	$13.5522 - 2.7321NPS + 1.4826t - 0.1483t^2$
Model 7 (treatment Gompertz w/o inter)	$0.3386t(0.1071Age\_group - 0.184Rad) * \exp(-0.3861 * Sex * t)$
Model 8 (NPS Gompertz w/o inter)	$0.2455t(0.275 - 0.026\text{SES} - 0.043\text{NPS}) * \exp\{(0.2538\text{SES} - 0.2455\text{Sex} - 0.1381\text{NPS})t\}$
Model 9 (NPS Gompertz w/ inter)	$0.2479t(0.275 - 0.026\text{SES} - 0.043\text{NPS}) * \exp\{(0.2538\text{SES} - 0.2479\text{Sex} - 0.1376\text{NPS})t\}$

with radiation starts to have lower composite IQ than patients without radiation and as time passing, the harm will increase. Also the harm of radiation is larger in female patients than male patients (see figure 8.6).

In figure 8.6, composite IQ of patients with and without radiation are compared in both male and female patient whose age is more than 7 years old at diagnosis with SES=3 .

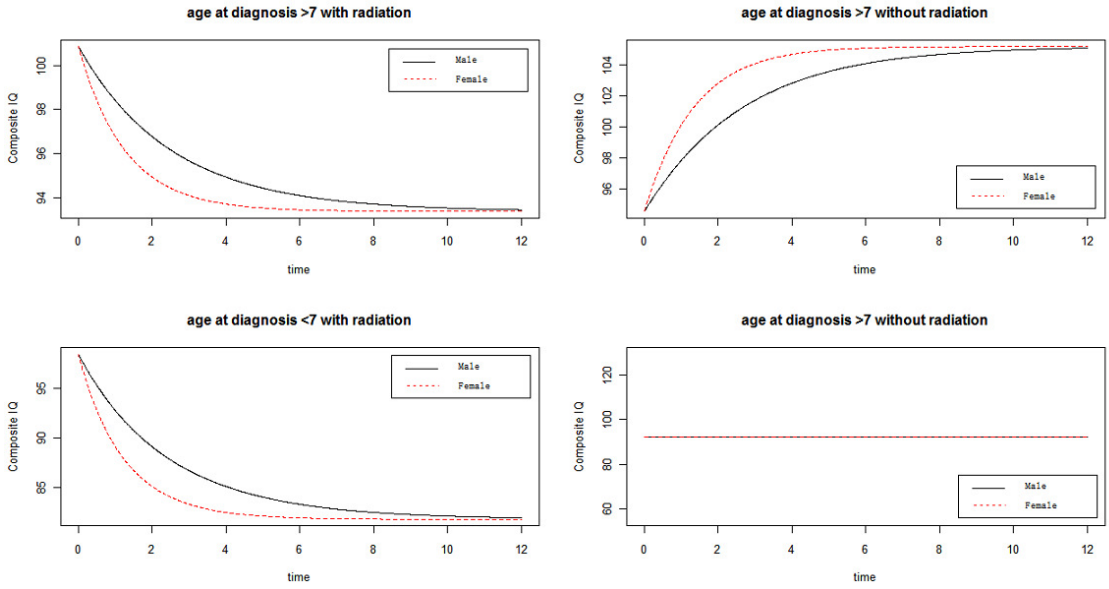

Figure 8.4 composite IQ comparison between male and female in different age group and radiation in model 7

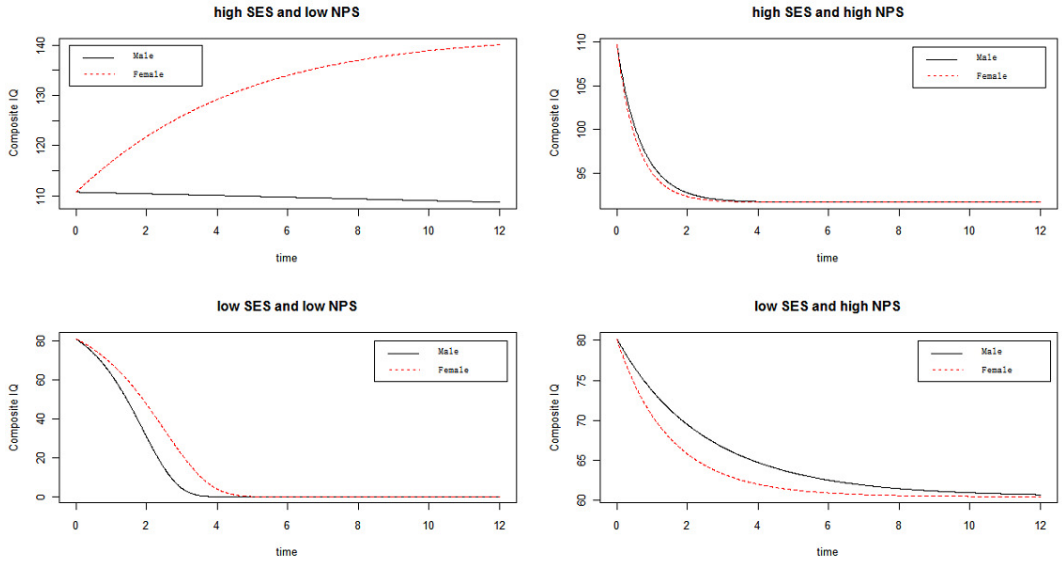

Figure 8.5 composite IQ comparison between male and female in different SES and NPS group in model 9

Table 8.7 Effect of radiation to composite IQ score

Model	$\frac{\partial IQ}{\partial Radiation}$ or $\frac{\partial \log IQ}{\partial Radiation}$
Model 1 (treatment linear w/o inter)	$-2.1324t$
Model 2 (treatment quadratic w/o inter)	$-2.0928t$
Model 7 (treatment Gompertz w/o inter)	$-0.1192+0.184\exp(-0.3861*Sex*t)$

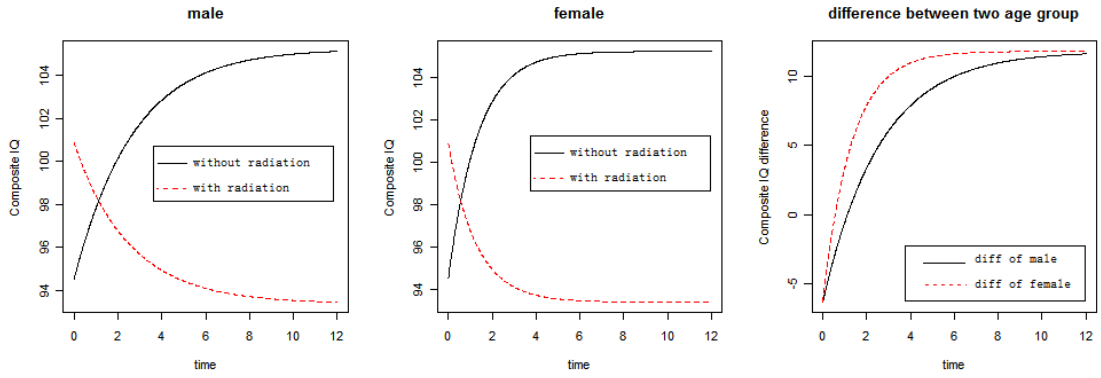

Figure 8.6 composite IQ comparison between radiation group in different sex in model 7

8.3.5 Effect of neurological predictor scale

Table 8.8 shows the effect of neurological predictor scale to composite IQ score. The hierarchical linear/quadratic models show that higher NPS will lead to lower composite IQ shortly after diagnosis and SES is higher (see figure 8.7), the harm of high NPS is larger. Also female will have more harm for high NPS than male (see figure 8.8). Furthermore, as time passing, the harm will increase. For Gompertz based hierarchical nonlinear model, the effect is not clear due to complex model structure.

In figure 8.7, composite IQ of patients with low NPS (NPS=0) and high NPS (NPS=10) are compared in both low SES (SES=5) and high SES (SES=1) female patient whose age is more than 7 years old at diagnosis.

In figure 8.8, composite IQ of patients with low NPS (NPS=0) and high NPS (NPS=10) are compared in both male and female patient whose age is more than 7 years old at diagnosis with SES=3 .

Table 8.8 Effect of neurological predictor scale to composite IQ score

Model	$\frac{\partial IQ}{\partial NPS}$ or $\frac{\partial \log IQ}{\partial NPS}$
Model 3 (NPS linear w/o inter)	$-1.3286-0.352t$
Model 4 (NPS linear w/ inter)	$0.748SES-2.5682Sex-0.36t$
Model 5 (NPS quadratic w/o inter)	$-1.147-0.3735t$
Model 6 (NPS quadratic w/ inter)	$0.866SES-2.7321Sex-0.4151t$
Model 8 (NPS Gompertz w/o inter)	- 0.044 + (0.043 + 0.0038t - 0.0036SES * t - 0.006NPS * t) * exp{(0.2538SES - 0.2455Sex - 0.1381NPS)t}
Model 9 (NPS Gompertz w/ inter)	- 0.044 + (0.043 + 0.0038t - 0.0036SES * t - 0.006NPS * t) * exp{(0.2538SES - 0.2479Sex - 0.1376NPS)t}

8.4 Difference between hierarchical linear/quadratic model and Gompertz based hierarchical nonlinear model in interpretation

Table 8.9 present some different interpretation between hierarchical linear/quadratic models and Gompertz based hierarchical nonlinear models. We can find that Gompertz models involve more information (interaction) than hierarchical linear/quadratic models. This

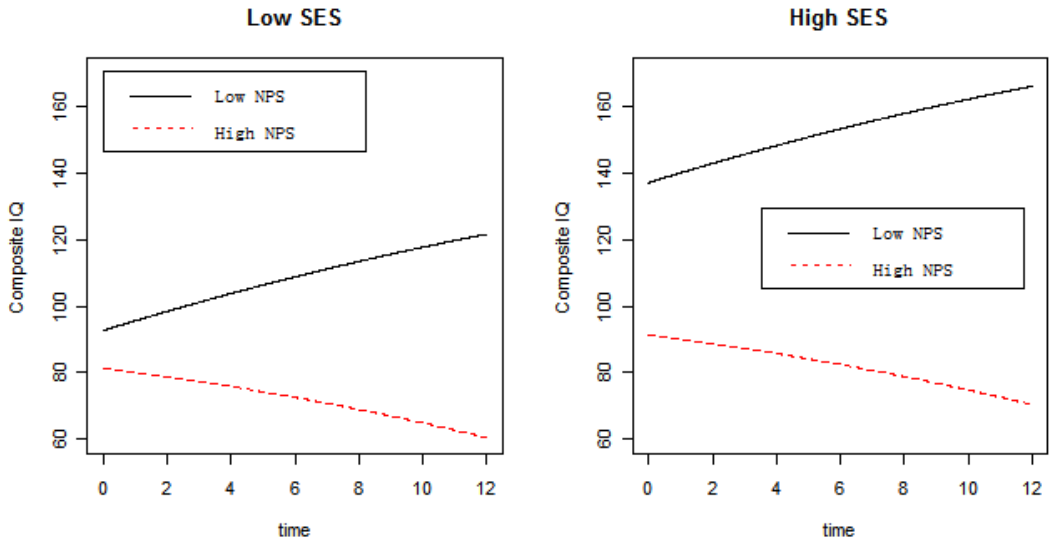

Figure 8.7 Composite IQ comparison between low and high NPS in low and high SES in model 6

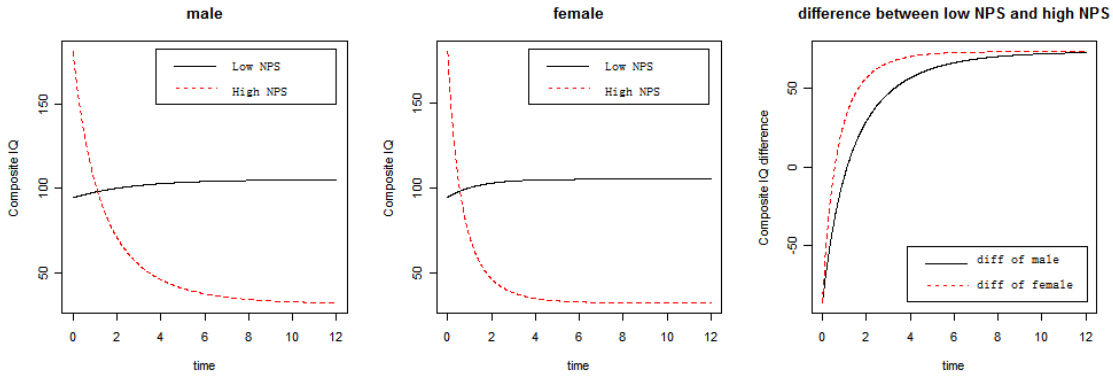

Figure 8.8 Composite IQ between low and high NPS in different sex in model 6

is the advantage of nonlinear models. But there is still some drawback: The information in the nonlinear model is not as easy to interpret as in hierarchical linear/quadratic models.

Table 8.9 Some different interpretation between hierarchical linear/quadratic model and Gompertz based nonlinear model

Variable	Hierarchical Linear/quadratic Model	Gompertz based nonlinear model
Age group	Simple positive effect	Positive effect involved Sex and SES into interpretation
SES	Positive effect involve NPS into interpretation	Simple positive effect or hard to interpret
Sex	Involve NPS into interpretation	Involve age group, radiation, SES and NPS into interpretation
Radiation	Simple negative effect	Negative effect involve Sex into interpretation
NPS	Negative effect Involve SES and Sex into interpretation	Hard to interpret

CHAPTER 9

Conclusion and future research

The new proposed VUS is an improved criterion for measuring the multi-class diagnostic accuracy by considering both sensitivity and specificity. And the estimation and inference is relatively easy to obtain. It is can be used as a tool in distinguishing the quality of given multi-classifier in diagnostic tests.

In the longitudinal study, the analysis using hierarchical linear/quadratic model and Gompertz based hierarchical nonlinear model shows the trend of pediatric brain-tumor patients' SB-IV test scores (specially, Composite IQ score) over time after diagnosis and the effect of each risk factors. The study validates the the finding of [25, 26] that the use of radiation therapy may cause the risk of suboptimal cognitive outcomes even though it can increase the survival rate. Also the model comparison and interpretation show the merit and drawback of Gompertz structure: Gompertz based hierarchical nonlinear model fits the data better but causes difficulty of model interpretation in some cases.

To construct models that fit the data better, some model transformation method can be applied and different model selection method may be employed. Also future study can be focused on how to conquer the difficulty of model interpretation for Gompertz based hierarchical nonlinear model.

REFERENCES

[1] C. Ferri J. Hernández-Orallo M.A. Salido (2003) Volume Under the ROC Surface for Multi-class Problems. Exact Computation and Evaluation of Approximations

[2] Christos T. Nakas and Constantin T. Yiannoutsos (2004) Ordered multiple-class ROC analysis with continuous measurements

[3] Eunkyung Cha (2005) The Volume Under The ROC Surface For Multi-ordered Classes Problem

[4] Donald Bamber (1975) The Area Above The Ordinal Dominance Graph and The Area Below the Receiver Operating Characteristic Graph

[5] Chenghong Shen (2006) Some significant results in the classification analysis of the spectroscopic evaluation of cervical cancer

[6] Hailun Wang (2007) Some conclusions of statistical analysis of the spectropscopic [sic] evaluation of cervical cancer

[7] E.L. Lehmann (1951] Elements of Large-Sample Theory 364-380

[8] Art B. Owen (2001) Empirical Likelihood

[9] Gengsheng Qin, Xiao-Hua Zhou (2005) Empirical Likelihood Inference for the Area Under the ROC Curve

[10] Thomas S. Ferguson (2005) U-statistics

[11] Abramowitz, Milton; Stegun, Irene A., eds. (1972), "§25.4, Integration", *Handbook of Mathematical Functions (with Formulas, Graphs, and Mathematical Tables)*, Dover, ISBN 978-0-486-61272-0

[12] Gil, Amparo; Segura, Javier; Temme, Nico M. (2007), "§5.3: Gauss quadrature", *Numerical Methods for Special Functions*, SIAM, ISBN 978-0-898716-34-4

[13] Kahaner, David; Moler, Cleve; Nash, Stephen (1989), *Numerical Methods and Software*, Prentice-Hall, ISBN 978-0-13-627258-8

[14] Press, William H.; Flannery, Brian P.; Teukolsky, Saul A.; Vetterling, William T. (1988), "§4.5: Gaussian Quadratures and Orthogonal Polynomials", *Numerical Recipes in C* (2nd ed.), Cambridge University Press, ISBN 978-0-521-43108-8

[15] Stoer, Josef; Bulirsch, Roland (2002), *Introduction to Numerical Analysis* (3rd ed.), Springer, ISBN 978-0-387-95452-3.

[16] Micklewright, J.L., King, T.Z., Morris, R.D., and Krawiecki, N. (in press) Quantifying Pediatric Neuro-oncology Risk Factors: Development of the Neurological Predictor Scale. *Journal of Child Neurology*.

[17] Francis, D., Fletcher, J., Stuebing, K., Davidson, K., and Thompson, N. (1991) Analysis of Change: Modeling Individual Growth. *Journal of Consulting and Clinical Psychology,* Vol. 59, No. 1, 27-37.

[18] Raudenbush, S., and Bryk, A.(2004) HLM6: Hierarchical Linear and Nonlinear Modeling. Chicago: Scientific Software International.

[19] Maas, C., and Snijders, T., (2003) The Multilevel Approach to Repeated Measure for Complete and Incomplete Data. *Quality and Quantity* 37: 71-89

[20] Laird A. K. (1964). "Dynamics of tumor growth". *Br J of Cancer* 18: 490-502.

[21] Steel, G.G. (1977). *Growth Kinetics of Tumors*. Oxford: Clarendon Press.

[22] Wheldon, T.E. (1988). *Mathematical Models in Cancer Research*. Bristol: Adam hilger.

[23] d'Onofrio A. (2005). "A general framework for modeling tumor-immune system competition and immunotherapy: Mathematical analysis and biomedical inferences". *Physica D* 208: 220-235.

[24] Eugene Demidenko (2004) Mixed Models Theory and Application, 291-296

[25] Ris, M. D., and Noll, R. B., (1994) Long-term neurobehavioral outcome in pediatric brain-tumor patients: review and methodological critique. *Journal of Clinical and Experimental Neuropsychology,* 16(1), 21-42

[26] Ris, M. D., Packer, R., Goldwein, J., Jones-Wallace, D., and Boyett, J. M., (2001) Intellectual outcome after reduced-dose radiation therapy plus adjuvant chemotherapy for medulloblastoma: A Children's Cancer Group study. *Journal of Clinical Oncology* Vol. 19, Issue 15 (August), 2001: 3470-3476

[27] Fuemmeler, B., Elkin, T., & Mullins, L. (2002). Survivors of childhood brain tumors: Behavioral, emotional and social adjustment. *Clinical Psychology Review*, 22, 547-585.

[28] Linet, M., Ries, L., Smith, M., Tarone, R., & Devesa, S. (1999). Cancer surveillance series: Recent trends in childhood cancer incidence and mortality in the United States. *Journal of the National Cancer Institute*, 91, 12, 1051-1058.

[29] Yueh-Yun Chi and Xiao-Hua Zhou (2006) Maximum Likelihood Inference on Three-way ROC Surfaces in the Presence of Differential Verifcation

[30] Ying Shen (2007) Individual Growth Models of Change in Peabody Picture Vocabulary Scores of Children Treated For Brain Tumors

[31] Gompertz, B. (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esc. F.R.S Phil. Trans. R. Soc. London 115, 513-585

[32] Perevozskaya I., Kuznetsova O.M. (2000). Modeling Longitudinal Growth Data and Growth Percentiles with Polynomial Gompertz Model in SAS Software. Proceedings of SUGI 25, paper 277-25.

[33] S. Zhang, O. Kuznetsova, J. Jacobs (1998), Use of a Gompertz curve to describe patterns of early growth in term and preterm infants, Human Biology Association Meeting, Salt Lake City, Uta.

[34] Davidian, M. and Giltinan, D.M. (1995) Nonlinear Models for Repeated Measurement Data, New York: Chapman & Hall.

[35] Pinheiro, J.C., and Bates, D.M. (1995) Model Building for Nonlinear Mixed-Effects Models. Technical Report 91.

[36] Huayan Chai (2007) Longitudinal Curves for Behaviors of Chilidren Diagnosed with a Brain Tumor

[37] H. Gerry Taylor and Jeremy Alden (1997) Age-related differences in outcomes following childhood brain insults: An introduction and overview. *Journal of the international Neuropsychological Society. 3. 555-567*

[38] Robert W. Butler (2007) Commentary: Lessons in Pediatric Neuropsycho-Oncology: What We're Learned Since Johnny Gunther. *Journal of Pediatric Psychology 32(9) pp. 1038-1039, 2007*

[39] M. Douglas Ris (2007) Lessons in Pediatric Neuropsycho-Oncology: What We Have Learned Since Johnny Gunther

[40] H. Gerry Talyor, Keith Owen Yeates, Shari L. Wade, Dennis Drotar, Terry Stancin, Nori Minich (2002) A Prospective Study of Short- and Long-Term Outcomes After Traumatic Brain Injury in Children: Behavior and Achievement. *Neuropsychology 2002, Vol. 16, No. 1, 15-27*

[41] Thorndike, R. L.., Hagen E.P. & Sattler, (1986) Measurement and Evaluation in Psychology and Education (4th Edition) Prentice Hall.

APPENDIX A

List of abbreviations

ROC: Receiver operating characteristic

AUC: The area under the ROC curve

STB-IV: Stanford-Binet Intelligence Scale-Fourth Edition

IQ: intelligence quotient

VUS: Volume under the surface

PRESS: Prediction Error Sum of Squares

HLM: Hierarchical Linear Model

WAPRESS: Weighted Adjusted PRESS

SES: Socioeconomic Status

NPS: Neurological Predictor Scale

EL: Empirical likelihood

ANOVA: Analysis of Variance

ANCOVA: analysis of covariance

ANCOVARC: analysis of covariance with reliability correction

ML: Maximum Likelihood

REML: Restricted/residual Maximum Likelihood

Compiq: Composite IQ

VRSAS: Verbal Reasoning SAS score

AVRSAS: Abstract/Visual Reasoning SAS score

QVRSAS: Quantitative Reasoning SAS score

STMSAS: Short-Term Memory SAS score

APPENDIX B

Analysis of four STB SAS scores

Verbal reasoning SAS score (VRSAS):

Model 1 (Hierarchical R&C Time):

$$VRSAS_{it} = (114.81 - 4.768SES_i + \mu_{0i}) + (-1.2286Rad_{it} + \mu_{1i})Time_{it} + \varepsilon_{it}$$

Model 2 (Gompertz R&C):

$$\log(VRSAS_{it}) = (4.0067 - 0.1479SES_i - 0.08304Chemo_{it} + \mu_{0i})$$
$$-(0.1696 - 0.09481SES_i - 0.05833Rad_{it} - 0.09453Chemo_{it}) *$$
$$e^{-(0.4889-0.1113SES_i+0.0733Rad_{it})Time_{it}} + \varepsilon_{it}$$

Model 3: (Hierarchical NPS Time)

$$VRSAS_{it} = (113.94 - 4.7469SES_i + \mu_{0i})$$
$$+(1.6503 - 0.3764NPS_{it} + \mu_{1i})Time_{it} + \varepsilon_{it}$$

Model 4 (Hierarchial NPS Time^2)

$$VRSAS_{it} = (114.91 - 4.7209SES_i + \mu_{0i})$$
$$+(-0.3129NPS_{it} + 1.2933Age\_group_i + \mu_{1i})Time_{it} + 0.09689Time_{it}^2 + \varepsilon_{it}$$

Model 5 (Gompertz NPS)

$$\log(VRSAS_{it}) = (4.8321 + 0.06155Age\_group_i - 0.03532SES_i$$
$$-0.03659NPS_{it} + \mu_{0i}) - (003086SES_i + 0.04946Sex_i - 0.03903NPS_{it}) *$$
$$e^{-(2.5193+0.07793SES_i-0.2807NPS_{it})Time_{it}} + \varepsilon_{it}$$

Model 6 (Hierarchical NPS Time with interaction)

$$VRSAS_{it} = (118.82 - 4.6533SES_i - 1.2867NPS_{it}$$
$$+0.9425NPS_{it}Age\_group_i + \mu_{0i}) + (-0.1182NPS_{it} + \mu_{1i})Time_{it} + \varepsilon_{it}$$

Model 7 (Gompertz NPS with interaction)

$$\log(VRSAS_{it}) = (4.8492 - 0.03515SES_i - 0.03988NPS_{it} + 0.01179Age\_group_i NPS_{it} + \mu_{0i}) - (0.03048SES_i + 0.05087Sex_i - 0.0392NPS_{it}) * e^{-(2.5932+0.08209SES_i - 0.2898NPS_{it})Time_{it}} + \varepsilon_{it}$$

Table B.1 WAPRESS for VRSAS

Without interaction	Hierarchical Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	20025.14	NA	18529.07
NPS	19675.25	19003.03	19924.20

With interaction	Hierarchical Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	NA	NA	NA
NPS	19355.16	NA	19635.13

3-level cutoff points: Low: <90; Median: 90-103; High: >103

4-level cutoff points: Low: <86; Median: 86-96; Median high: 97-107; High: >107

Table B.2 3-level VUS and transformed VUS for VRSAS models

Without interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	0.3264778 (0.7095111) * (0.7482239) **	NA	0.3405685 (0.7268535) (0.765421)
Neurological Predictor Scale	0.3198148 (7013105) (0.7399628)	0.3297046 (0.7134826) (0.7521944)	0.3142836 (0.6945029) (0.7330395)

With interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	NA	NA	NA
Neurological Predictor Scale	0.3333403 (0.7179573) (0.7566447)	NA	0.3142469 (0.6944577) (0.7329933)

(* linear transformed VUS **quadratic transformed VUS)

Table B.3 4-level VUS and transformed VUS for VRSAS models

Without interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	0.0801443 (0.6343053) * (0.711651) **	NA	0.0880193 (0.6544653) (0.7341624)
Neurological Predictor Scale	0.0807018 (0.6357324) (0.7132919)	0.0822927 (0.6398052) (0.7179335)	0.0827538 (0.6409856) (0.7192676)

With interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	NA	NA	NA
Neurological Predictor Scale	0.0833406 (0.6424879) (0.7209583)	NA	0.0824069 (0.6400975) (0.7182643)

(* linear transformed VUS **cubic transformed VUS)

Model Interpretation for VRSAS models:

SES:

All hierarchical models indicate lower SES score will lead to lower verbal reasoning SAS score. For all the Gompertz models, the trend is not clear because of the more complicated model structure.

Radiation:

Hierarchical model indicates radiation is harmful to verbal reasoning SAS score and when time is passing, the harm will increase. For Gompertz model, the trend is not clear because of the more complicated model structure.

Chemotherapy:

For hierarchical model, chemo is not significant. And for Gompertz model, Chemo is significant but the trend is not clear because of the more complicated model structure.

Age_group:

For hierarchical models, age_group is significant in NPS model and indicate that the individual who is 7 years older at diagnosis will have higher verbal reasoning SAS score and the advantage of older age will increase when time is passing and also the advantage of older age is larger when the NPS is larger. For Gompertz model, the same conclusion will be made as hierarchical model.

Sex:

For all the hierarchical models, sex is not significant. But for Gompertz NPS models, sex is significant and we can conclude male will have higher verbal reasoning SAS score and for lower SES and lower NPS, the advantage of gender is larger.

NPS:

For all the hierarchical models, higher NPS will lead lower verbal reasoning SAS score and when time is passing, the harm will increase. For all the Gompertz models, the trend is not clear because of the more complicated model structure. Hierarchical model with interaction indicates that for patient whose age is less than 7 years old at diagnosis, the harm of high NPS is larger.

Abstract/visual reasoning SAS score (AVRSAS):

Model 1 (Hierarchical R&C Time):

$$AVRSAS_{it} = (112.20 + 9.5203\,Age\_group_i - 7.5410\,SES_i + \mu_{0i}) + (-2.2262\,Rad_{ti} + 0.7082\,SES_i + \mu_{1i})Time_{it} + \varepsilon_{it}$$

Model 2 (Gompertz R&C):

$$\log(AVRSAS_{it}) = (4.8207 + 0.09876\,Age\_group_i - 0.05444\,Sex_i - 0.2038\,Rad_{it} + \mu_{0i}) - (0.1998 - 0.2205\,Rad_{it})e^{-0.1983\,Sex_i Time_{it}} + \varepsilon_{it}$$

Model 3: (Hierarchical NPS Time)

$$AVRSAS_{it} = (114.06 - 5.1356\,SES_i + 9.2328\,Age\_group_i - 1.8857\,NPS_{it} + \mu_{0i}) + (3.2861 - 0.3376\,NPS_{it} + \mu_{1i})Time_{it} + \varepsilon_{it}$$

Model 4 (Gompertz NPS)

$$\log(AVRSAS_{it}) = (4.8015 + 0.09837\,Age\_group_i - 0.05716\,SES_i + \mu_{0i}) - 0.02870\,NPS_{ti}\,e^{-0.06281\,Time_{it}} + \varepsilon_{it}$$

Model 5 (Hierarchical NPS Time with interaction)

$$AVRSAS_{it} = (87.3320 + 19.5394\,Sex_i + 9.3322\,Age\_group_i - 1.5480\,NPS_{it}\,Sex_i - 3.4872\,SES_i\,Sex_i + \mu_{0i}) + (3.2786 - 0.3249\,NPS_{it} + \mu_{1i})Time_{it} + \varepsilon_{it}$$

Table B.4 WAPRESS for AVRSAS

Without interaction	Hierarchical Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	27252.15	NA	25222.99
NPS	26276.53	NA	21526.71

With interaction	Hierarchical Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	NA	NA	NA
NPS	26171.48	NA	NA

3-level cutoff points: Low: <87; Median: 87-102; High: >102

4-level cutoff points: Low: <83; Median: 83-94; Median high: 95-107; High: >107

Table B.5 3-level VUS and transformed VUS for AVRSAS models

Without interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	0.2925914 (0.6678048) * (0.7052808) **	NA	0.2992109 (0.675951) (0.7138575)
Neurological Predictor Scale	0.3188824 (0.700163) (0.7388)	NA	0.3204386 (0.7020782) (0.7407399)

With interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	NA	NA	NA
Neurological Predictor Scale	0.3251453 (0.7078712) (0.7465787)	NA	NA

(* linear transformed VUS **quadratic transformed VUS)

Table B.6 4-level VUS and transformed VUS for AVRSAS models

Without interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	0.07982 (0.6334751) * (0.7106929) **	NA	0.0795692 (0.632833) (0.7099501)
Neurological Predictor Scale	0.0770225 (0.6263135) (0.702318)	NA	0.0863848 (0.650281) (0.729604)

With interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	NA	NA	NA
Neurological Predictor Scale	0.0809446 (0.6363541) (0.7140043)	NA	NA

(* linear transformed VUS **cubic transformed VUS)

Model Interpretation for AVRSAS models:

SES:

For hierarchical models, lower SES score will lead to lower Abstract/Visual Reasoning SAS score and when time is passing, the harm will increase. Also female individuals have more harm.

The Gompertz NPS model indicates lower SES will lead to lower Abstract/Visual Reasoning SAS score.

Radiation:

Hierarchical model indicates radiation is harmful to Abstract/Visual Reasoning SAS score and when time is passing, the harm will increase. Gompertz model indicates for male, the harm of radiation to Abstract/Visual Reasoning SAS score get start after 0.4 years and when time is passing, the harm will increase, and for female, the harm of radiation to Abstract/Visual Reasoning SAS score get start after 0.2 years and when time is passing, the harm will increase.

Chemotherapy:

Chemo is not significant for all models.

Age_group:

All indicates that the individual who is 7 years older at diagnosis will have higher Abstract/Visual Reasoning SAS score

Sex:

For the hierarchical models, sex is only significant in NPS model with interaction and indicates that female have higher Abstract/Visual Reasoning SAS score except for high NPS in low SES individuals. For Gompertz R&C model, the trend is not clear because of the more complicated model structure.

NPS:

For all models, higher NPS will lead lower Abstract/Visual Reasoning SAS score and when time is passing, the harm will increase. And hierarchical model with interaction indicates that for female, the harm of high NPS is larger.

Quantitative reasoning SAS score (QRSAS):

Model 1 (Hierarchical R&C Time):

$$QRSAS_{it} = (115.48 - 5.4432SES_i + \mu_{0i}) + (-1.3953Rad_{it} - 0.5346Sex_i + \mu_{1i})Time_{it} + \varepsilon_{it}$$

Model 2 (Gompertz R&C):

$$\log(QRSAS_{it}) = (4.6482 - 0.08216SES_i - 0.07408Chemo_{it} + \mu_{0i}) - (-0.09956Sex_i - 0.1336Rad_{it})e^{-(-0.05564SES_i + 0.1253Sex_i + 0.1253Rad_{it})Time_{it}} + \varepsilon_{it}$$

Model 3 (Gompertz R&C with interaction):

$$\log(QRSAS_{it}) = (4.6027 - 0.06187SES_i - 0.05708Chemo_{it} - 0.1214Rad_{it}Sex_i + \mu_{0i}) - (-0.09094\,Sex_i - 0.2678Rad_{it})e^{-0.105Rad_{it}Time_{it}} + \varepsilon_{it}$$

Model 4: (Hierarchical NPS Time)

$$QRSAS_{it} = (116.37 - 5.7879SES_i + \mu_{0i}) + (-0.2624NPS_{it} + \mu_{1i})Time_{it} + \varepsilon_{it}$$

Model 5 (Gompertz NPS)

$$\log(QRSAS_{it}) = (4.5379 - 0.01929NPS_{it} + \mu_{0i}) - (0.02340SES_i - 0.1428Sex_i)e^{-(-0.1311 - +.1128SES_i)Time_{it}} + \varepsilon_{it}$$

Model 6 (Hierarchical NPS Time with interaction)

$$QRSAS_{it} = (113.18 + 15.7174Sex_i + 5.1958Age\_group_i - 11.7858SES_i - 2.8613NPS_{it}Sex_i + 1.0849SES_iNPS_{it} + \mu_{0i}) + (-0.2265NPS_{it} + \mu_{1i})Time_{it} + \varepsilon_{it}$$

Model 7 (Gompertz NPS with interaction)

$$\log(QRSAS_{it}) = (4.5075 - 0.03258SES_iSex_i - 0.01272Sex_iNPS_{it} + \mu_{0i}) - (-0.2154Sex_i)e^{-(0.01948SES_i)Time_{it}} + \varepsilon_{it}$$

Table B.7 WAPRESS for QRSAS

Without interaction	Hierarchical Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	20475.59	NA	20995.87
NPS	19995.72	NA	23071.07

With interaction	Hierarchical Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	NA	NA	20313.56
NPS	18953.44	NA	20412.90

3-level cutoff points: Low: <86; Median: 86-100; High: >100

4-level cutoff points: Low: <82; Median: 82-93; Median high: 94-104; High: >104

Table B.8 3-level VUS and transformed VUS for QRSAS models

Without interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	0.326183 (0.7091483) * (0.7478602) **	NA	0.3327869 (0.7172762) (0.7559689)
Neurological Predictor Scale	0.3271052 (0.7102833) (0.7489974)	NA	0.3025735 (0.6800904) (0.7181779)

With interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	NA	NA	0.3428007 (0.7296008) (0.7681123)
Neurological Predictor Scale	0.3448607 (0.7321363) (0.7705885)	NA	0.3255986 (0.7084291) (0.7471388)

(* linear transformed VUS **quadratic transformed VUS)

Table B.9 4-level VUS and transformed VUS for QRSAS models

Without interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	0.081203 (0.6370154) * (0.7147606) **	NA	0.075581 (0.6226232) (0.697924)
Neurological Predictor Scale	0.0755356 (0.6225069) (0.6977845)	NA	0.0702476 (0.6089697) (0.6811612)

With interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	NA	NA	0.078268 (0.6295019) (0.7060715)
Neurological Predictor Scale	0.0787384 (0.630706) (0.7074785)	NA	0.0743405 (0.6194474) (0.6940971)

(* linear transformed VUS **cubic transformed VUS)

Model Interpretation for QRSAS models:

SES:

All hierarchical models indicate lower SES score will lead to lower Quantitative Reasoning SAS score. All the Gompertz models except R&C model indicate lower SES score will lead to lower Quantitative Reasoning SAS score. For Gompertz R&C model, the trend is not clear because of the more complicated model structure. Hierarchical NPS model with interaction indicates that the lower of NPS, the harm of low SES will be larger.

Radiation:

Hierarchical model indicates radiation is harmful to quantitative reasoning SAS score and when time is passing, the harm will increase. For Gompertz model, the trend is not clear because of the more complicated model structure.

Chemotherapy:

For hierarchical model, chemo is not significant. And Gompertz model indicates that chemo is harmful to Quantitative Reasoning SAS score.

Age_group:

For hierarchical models, age_group is significant in NPS model and indicate that the individual who is 7 years older at diagnosis will have higher Quantitative Reasoning SAS score.

Sex:

Hierarchical R&C model indicates male has higher Quantitative Reasoning SAS score than female and when time is passing, the advantage of male is larger. And Gompertz R&C model with interaction indicates for individual without radiation, female will have higher Quantitative Reasoning SAS score but for individual with radiation, male will have higher Quantitative Reasoning SAS score and when time is passing, the difference will be larger. Hierarchical NPS model with interaction indicates female individuals have higher Quantitative Reasoning SAS score when NPS is smaller than 6 and male individuals have higher Quantitative Reasoning SAS score when NPS is larger or equal to 6.

NPS:

For all the models, higher NPS will lead lower quantitative reasoning SAS score. And in hierarchical model with interaction we can find that for female, the harm of high NPS is larger than male and for higher SES, the harm of high NPS is larger.

Short-term memory SAS score (STMSAS):

Model 1 (Hierarchical R&C Time):

$$STMSAS_{it} = (107.87 + 9.1240\,Age\_group_i - 5.8583\,SES_i + \mu_{0i}) + (-1.9732\,Rad_{it} + 0.8134\,Sex_i + \mu_{1i})Time_{it} + \varepsilon_{it}$$

Model 2 (Hierarchical R&C Time^2):

$$STMSAS_{it} = (110.43 - 5.6126\ SES_i + \mu_{0i}) + (-5.6126\ Rad_{it} + 0.6098 Sex_i + 4.4244 Age\_group_i + \mu_{1i}) Time_{it} + (-0.5184 Age\_group_i + 0.1281 Rad_{it}) Time_{it}^{\ 2} + \varepsilon_{it}$$

Model 3 (Hierarchical R&C Time with interaction):

$$STMSAS_{it} = (112.18 - 7.2173 SES_i + 0.8447 Age\_group_i SES_i + \mu_{0i}) + (-1.9712 Rad_{it} - 1.9712 Sex_i + \mu_{1i}) Time_{it} + \varepsilon_{it}$$

Model 4 (Gompertz R&C):

$$\log(STMSAS_{it}) = (4.6671 + 0.1231 Age\_group_i - 0.05023 SES_i - 0.09455 Rad_{it} + \mu_{0i}) - (0.1315 Age\_group_i + 0.0555 Rad_{it} + \mu_{1i}) e^{-(0.7875 SES_i - 0.7412 Chemo_{it}) Time_{it}} + \varepsilon_{it}$$

Model 5 (Gompertz R&C with interaction):

$$\log(STMSAS_{it}) = (4.7236 - 0.06876 SES_i - 0.09188 Rad_{it} + 0.009982 Age\_group_i SES_i + \mu_{0i}) - (0.1266 Age\_group_i - 0.1857 Rad_{it} + \mu_{1i}) * e^{-(0.7894 SES_i - 0.7364 Chemo_{it}) Time_{it}} + \varepsilon_{it}$$

Model 6: (Hierarchical NPS Time)

$$STMSAS_{it} = (115.47 + \beta_{01} SES_i + 9.0006 Age\_group_i - 1.6599 NPS_{it} + \mu_{0i}) + (-0.2295 NPS_{it} + 1.1119\ Sex_i + \mu_{1i}) Time_{it} + \varepsilon_{it}$$

Model 7: (Hierarchical NPS Time with interaction)

$$STMSAS_{it} = (140.70 - 13.9719 SES_i - 5.6364\ NPS_{it} + 2.9496 Age\_group_i SES_i + 1.2909 NPS_{it} SES_i + \mu_{0i}) + (-0.2316 NPS_{it} + 1.1306 Sex_i + \mu_{1i}) Time_{it} + \varepsilon_{it}$$

Model 8 (Gompertz NPS)

$$\log(STMSAS_{it}) = (4.7747 + 0.1227 Age\_group_i - 0.05659 SES_i - 0.02441 NPS_{it} + \mu_{0i}) - (0.1202 Age\_group_i - 0.01355 NPS_{it} + \mu_{1i}) e^{-0.5167 SES_i Time_{it}} + \varepsilon_{it}$$

Model 9 (Gompertz NPS with interaction)

$$\log(STMSAS_{it}) = (5.0598 - 0.1536\,SES_i - 0.06733\,NPS_{it} + 0.04 Age\_group_i SES_i + 0.01431 NPS_{it} SES_i + \mu_{0i}) - (0.1156 Age\_group_i + -0.01487 NPS_{it} + \mu_{1i}) * e^{-0.5377 SES_i Time_{it}} + \varepsilon_{it}$$

Table B.10 WAPRESS for STMSAS

Without interaction	Hierarchical Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	24155.35	23977.59	24784.84
NPS	24304.18	NA	23672.98

With interaction	Hierarchical Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	23979.33	NA	24867.42
NPS	23845.62	NA	23359.86

3-level cutoff points: Low: <85; Median: 85-100; High: >100

4-level cutoff points: Low: <81; Median: 81-92; Median high: 93-105; High: >105

Table B.11 3-level VUS and transformed VUS for STMSAS models

Without interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	0.335184 (0.7202265) * (0.7588922) **	0.3326626 (0.7171232) (0.755817)	0.2696163 (0.6395278) (0.6747334)
Neurological Predictor Scale	0.3413209 (0.7277796) (0.7663292)	NA	0.3457757 (0.7332624) (0.7716858)

With interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	0.3349308 (0.7199148) (0.7585839)	NA	0.325498 (0.7083052) (0.7470144)
Neurological Predictor Scale	0.3183853 (0.6995512) (0.7381794)	NA	0.3419241 (0.728522) (0.7670566)

(* linear transformed VUS **quadratic transformed VUS)

Table B.12 4-level VUS and transformed VUS for STMSAS models

Without interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	0.0753122 (0.6219351) * (0.6970984) **	0.0736045 (0.6175634) (0.6918068)	0.059655 (0.5818526) (0.6451647)
Neurological Predictor Scale	0.0755203 (0.6224677) (0.6977376)	NA	0.0765059 (0.624991) (0.70075)

With interaction	Hierarchical linear/quadratic Model		Gompertz Model
	Time	Time^2	
Radiation & Chemo	0.073019 (0.6160644) (0.6899735)	NA	0.0646774 (0.5947101) (0.6627219)
Neurological Predictor Scale	0.0694119 (0.6068302) (0.6784583)	NA	0.0836986 (0.6434041) (0.7219857)

(* linear transformed VUS **cubic transformed VUS)

Model Interpretation for STMSAS models:

SES:

All hierarchical models indicate lower SES score will lead to lower Short-term Memory SAS score. For the Gompertz models the trend is not clear because of the more complicated model structure.

Radiation:

Hierarchical model indicates radiation is harmful to Short-term Memory and also when time is passing, the harm will increase. For the Gompertz models the trend is not clear because of the more complicated model structure.

Chemotherapy:

For all hierarchical models, chemo is not significant. For the Gompertz models the chemo is significant but the trend is not clear because of the more complicated model structure.

Age_group:

Hierarchical model indicates that the individual who is 7 years older at diagnosis will have higher Short-term Memory SAS score. Hierarchical model with interaction indicates that for lower SES, the advantage of old diagnosis age is larger. Gompertz model indicates that for lower SES, the advantage of old diagnosis age is larger and when time is passing, the advantage of old diagnosis age is larger.

Sex:

Hierarchical R&C model indicates female has higher Short-term Memory SAS score than male and as time passing, the advantage will be increasing. And for all Gompertz models, sex is not significant.

NPS:

For all the hierarchical models, higher NPS will lead lower Short-term Memory SAS score and as time passing, the harm of high NPS is larger. Gompertz model indicates that for lower SES, the harm of high NPS is larger and when time is passing, the harm of high NPS is larger.

APPENDIX C

Selected SAS and R code

VUS calculation for some models in cervical cancer diagnosis example

```
libname After 'H:\SAS Code for Algorithm\sas data';
data combine;
set after.clean510;
run;
data combine;
set combine;
newpap=preferredpap;
z1=p10ra32/p50ra40;
z2=p10ra68/p50ra44;
z3=p10ra67/p75ra42;
z4=p25ra24/p75ra29;
z5=p25ra67/p75ra42;
z6=p25ra67/p75ra44;
z7=p50ra67/p75ra42;
z8=p75ra42/p90ra67;
z9=p75ra43/p90ra67;

r1=p10ra4/p10ra29;
r2=p10ra5/p10ra22;
r3=p10ra30/p10ra48;
r4=p25ra30/p25ra45;
r5=p50ra33/p50ra54;
r6=p75ra31/p75ra50;
r7=p75ra64/p75ra75;
r8=p90ra6/p90ra29;
r9=p90ra10/p90ra29;
r10=p90ra32/p90ra51;
r11=p90ra66/p90ra75;

if newpap=2.8 then newpap=3.5;
if newpap=3.2 then newpap=2;
run;

data combine;
set combine;
if whole1=2.5 then delete;
else if whole1<2 then y=0;
else if whole1>=3 then y=2;
else y=1;
run;

proc pls data =combine;
      model high =p25ra1 p25ra3 p25ra5 p25ra30-p25ra31 p75ra25-p75ra29 z1-z9
newpap;
      output out=pred pred=pred;
run;

ods listing close;
```

```
ods html close;
ods trace off;

ods output Association=auc;
proc logistic data=pred;
model high=pred;
run;

data auc;
set auc;
keep label2 nvalue2;
if _n_=4;
run;

data auc;
set auc;
var="p25ra1 p25ra3 p25ra5 p25ra30-p25ra31 p75ra25-p75ra29 z1-z9 newpap";
type='    training    ';
vus=0;
run;

data output_auc;
set auc;
if _n_=1 then delete;
run;

%macro sspec(datain=, applydata=, var=, scale=);

*********************************
find auc
*********************************;

data cin1out; set &datain; if whole1 not in (2.5);whole = (whole1 > 2); run;
data apply_cin1out; set &applydata; if whole1 not in (2.5); whole = (whole1 >
2); run;


proc pls data = cin1out;
      model whole= &var;
      output out=pred pred=pred;
run;

ods listing close;
ods html close;
ods trace off;

ods output Association=auc;
proc logistic data=pred;
model high=pred;
run;

data auc;
set auc;
keep label2 nvalue2;
if _n_=4;
run;

data auc;
set auc;
var="&var";
```

```
type='training';
run;

*************************
find VUS
*************************;

proc pls data=combine;
model y=&var;
output out=out predicted=pred;
run;
data out1;
set out;
if y=0;
keep pred;
run;
data out2;
set out;
if y=1;
keep pred;
run;
data out3;
set out;
if y=2;
keep pred;
run;
proc iml;
use out1;
read all into x;
close out1;
use out2;
read all into y;
close out2;
use out3;
read all into z;
close out3;
n1=nrow(x);n2=nrow(y);n3=nrow(z);
s=0;
do i=1 to n1;
do j=1 to n2;
do k=1 to n3;
a=0;b=0;c=0;
if x[i,]<=y[j,] then a=1;
if y[j,]<=z[k,] then b=1;
if x[i,]<=z[k,] then c=1;
s=s+c*(a/3+1/3)*(b/3+1/3);
end;
end;
end;
vus=s/(n1*n2*n3);
create vus var{vus};
append;
close vus;
quit;

data auc;
merge auc vus;
run;

data output_auc;
```

```
set output_auc auc;
run;


%mend;

%macro sspec_10x(datain=, applydata=, var=, scale=);

**********************
find auc
**********************;

data cin1out; set &datain; if whole1 not in (2.5);whole = (whole1 > 2); run;
data apply_cin1out; set &applydata; if whole1 not in (2.5); whole = (whole1 >
2); run;


proc pls data = cin1out cv=split(10);
      model whole = &var;
      output out=pred pred=pred;
run;

ods listing close;
ods html close;
ods trace off;

ods output Association=auc;
proc logistic data=pred;
model high=pred;
run;

data auc;
set auc;
keep label2 nvalue2;
if _n_=4;
run;

data auc;
set auc;
var="&var";
type='10x validation';
run;

*************************
find VUS
*************************;

proc pls data=combine cv=split(10);
model y=&var;
output out=out predicted=pred;
run;
data out1;
set out;
if y=0;
keep pred;
run;
data out2;
set out;
if y=1;
keep pred;
```

```
run;
data out3;
set out;
if y=2;
keep pred;
run;
proc iml;
use out1;
read all into x;
close out1;
use out2;
read all into y;
close out2;
use out3;
read all into z;
close out3;
n1=nrow(x);n2=nrow(y);n3=nrow(z);
s=0;
do i=1 to n1;
do j=1 to n2;
do k=1 to n3;
a=0;b=0;c=0;
if x[i,]<=y[j,] then a=1;
if y[j,]<=z[k,] then b=1;
if x[i,]<=z[k,] then c=1;
s=s+c*(a/3+1/3)*(b/3+1/3);
end;
end;
end;
vus=s/(n1*n2*n3);
create vus var{vus};
append;
close vus;
quit;

data auc;
merge auc vus;
run;

data output_auc;
set output_auc auc;
run;
%mend;

/*model 1.03*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra78, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra78,scale=0);
/*model 1.04*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra19 p25ra30-p25ra38
p25ra59-p25ra68 p75ra20-p75ra29 p75ra39-p75ra58 p75ra69-p75ra78, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra19 p25ra30-
p25ra38 p25ra59-p25ra68 p75ra20-p75ra29 p75ra39-p75ra58 p75ra69-
p75ra78,scale=0);
/*model 2.0*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra31, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra31,scale=0);
/*model 2.1*/
```

```
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 p75ra57 p75ra58, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 p75ra57 p75ra58,scale=0);
/*model 2.11*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 p75ra57 p75ra58 sb1-sb11, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 p75ra57 p75ra58 sb1-sb11,scale=0);
/*model 2.12*/
 %sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 p75ra57 p75ra58 sbr1-sbr15, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 p75ra57 p75ra58 sbr1-sbr15,scale=0);
/*model 2.13*/
 %sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 p75ra57 p75ra58 sbr16-sbr30, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 p75ra57 p75ra58 sbr16-sbr30,scale=0);
/*model 2.14*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 p75ra57 p75ra58 sbr31-sbr45, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 p75ra57 p75ra58 sbr31-sbr45,scale=0);
/*model 2.15*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 p75ra57 p75ra58 sbr46-sbr60, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 p75ra57 p75ra58 sbr46-sbr60,scale=0);
/*model 2.20*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 p75ra57 p75ra58 sb12-sb22, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 p75ra57 p75ra58 sb12-sb22,scale=0);
/*model 2.3*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 p75ra57 p75ra58 r1-r11, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 p75ra57 p75ra58 r1-r11,scale=0);
/*model 2.31*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 p75ra57 p75ra58 r1-r11, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 p75ra57 p75ra58 r1-r11,scale=0);
/*model 2.32*/
%sspec(datain=combine, applydata=combine, var=p25ra1 p25ra3 p25ra5 p25ra30
p25ra32 p75ra25 p75ra27 p75ra29 p75ra57 p75ra58 r1-r8 r10 r11, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1 p25ra3 p25ra5 p25ra30
p25ra32 p75ra25 p75ra27 p75ra29 p75ra57 p75ra58 r1-r8 r10 r11,scale=0);
/*model 2.33*/
%sspec(datain=combine, applydata=combine, var=p25ra3 p25ra31 p75ra27 p75ra29
p75ra57 r2 r4 r5 r7 r8 r10 r11, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra3 p25ra31 p75ra27
p75ra29 p75ra57 r2 r4 r5 r7 r8 r10 r11,scale=0);
/*model 2.35*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p75ra25-p75ra29
p75ra57-p75ra58 r1-r11 gm1, scale=0);
```

```
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p75ra25-
p75ra29 p75ra57-p75ra58 r1-r11 gm1,scale=0);
/*model 2.36*/
 %sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p75ra25-p75ra29
r1-r11 gm1 gm3, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p75ra25-
p75ra29 r1-r11 gm1 gm3,scale=0);
/*model 2.37*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 r2 r4 r5 r6 r8 r9 r11, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 r2 r4 r5 r6 r8 r9 r11,scale=0);
/*model 2.39*/
%sspec(datain=combine, applydata=combine, var=p25ra4 p25ra5 p25ra30 p25ra31
p25ra32 p75ra28 p75ra29 p75ra57 p75ra58 r1-r11, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra4 p25ra5 p25ra30
p25ra31 p25ra32 p75ra28 p75ra29 p75ra57 p75ra58 r1-r11,scale=0);
/*model 2.40*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 p75ra58 r1 r2 r4 r6 r8 r9 r11, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 p75ra58 r1 r2 r4 r6 r8 r9 r11,scale=0);
/*model 2.41*/
%sspec(datain=combine, applydata=combine, var=p10ra30 p25ra32 dmr1-dmr11,
scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p10ra30 p25ra32 dmr1-
dmr11,scale=0);
/*model 2.42*/
%sspec(datain=combine, applydata=combine, var=dmr1-dmr18, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=dmr1-dmr18,scale=0);
/*model 2.43*/
%sspec(datain=combine, applydata=combine, var=dmr1-dmr18 d1-d8, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=dmr1-dmr18 d1-d8,scale=0);
/*model 2.44*/
 %sspec(datain=combine, applydata=combine, var=p25ra1-p25ra4 p25ra30-p25ra32
p75ra25-p75ra29 dmr1-dmr3 dmr5 dmr6 dmr9 dmr11 dmr14 dmr15 d1 d3 d4 d7,
scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra4 p25ra30-
p25ra32 p75ra25-p75ra29 dmr1-dmr3 dmr5 dmr6 dmr9 dmr11 dmr14 dmr15 d1 d3 d4
d7,scale=0);
/*model 2.45*/
 %sspec(datain=combine, applydata=combine, var=p10m1 p25m1 p25m2 p50m1 p75m1
p75m2 p90m1 p10_75m1 p10_75m2 p10_75m3 p10_75m4 p10_90m1 p10_90m2 p10_90m3
p25_75m1 p25_75m2 p25_75m3 p25_75m4 p25_75m5 p25_90m1 p25_90m2 p25_90m3,
scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p10m1 p25m1 p25m2 p50m1
p75m1 p75m2 p90m1 p10_75m1 p10_75m2 p10_75m3 p10_75m4 p10_90m1 p10_90m2
p10_90m3 p25_75m1 p25_75m2 p25_75m3 p25_75m4 p25_75m5 p25_90m1 p25_90m2
p25_90m3,scale=0);
/*model 2.46*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 p75ra57-p75ra58 p10m1 p25m1 p25m2 p50m1 p75m1 p75m2 p90m1
p10_75m1 p10_75m2 p10_75m3 p10_75m4 p10_90m1 p10_90m2 p10_90m3 p25_75m1
p25_75m2 p25_75m3 p25_75m4 p25_75m5 p25_90m1 p25_90m2 p25_90m3, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 p75ra57-p75ra58 p10m1 p25m1 p25m2 p50m1 p75m1 p75m2
p90m1 p10_75m1 p10_75m2 p10_75m3 p10_75m4 p10_90m1 p10_90m2 p10_90m3 p25_75m1
p25_75m2 p25_75m3 p25_75m4 p25_75m5 p25_90m1 p25_90m2 p25_90m3,scale=0);
```

```
/*model 2.47*/
%sspec(datain=combine, applydata=combine, var=p10m1 p25m1 p25m2 p75m2 p90m1
p10_75m2 p10_90m1 p25_75m2 p25_75m4 p25_75m5 p25_90m3, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p10m1 p25m1 p25m2 p75m2
p90m1 p10_75m2 p10_90m1 p25_75m2 p25_75m4 p25_75m5 p25_90m3,scale=0);
/*model 2.48*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra29 p75ra57-p75ra58 p10m1 p25m1 p25m2 p75m2 p90m1 p10_75m2
p10_90m1 p25_75m2 p25_75m4 p25_75m5 p25_90m3, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra29 p75ra57-p75ra58 p10m1 p25m1 p25m2 p75m2 p90m1 p10_75m2
p10_90m1 p25_75m2 p25_75m4 p25_75m5 p25_90m3,scale=0);
/*model 2.52*/
%sspec(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-p25ra32
p75ra25-p75ra28 p75ra57 y1-y10, scale=0);
%sspec_10x(datain=combine, applydata=combine, var=p25ra1-p25ra5 p25ra30-
p25ra32 p75ra25-p75ra28 p75ra57 y1-y10,scale=0);
```

Theoretical VUS and U-statistics, bootstrap simulation for normal mu1=1 mu2=2 mu3=3, and sigma1=sigma2=sigma3=0.5 and sample size 40. For other cases, the code is similar.

```
proc iml;
a1=0.4;a2=0.2;a3=1-a1-a2;
n1=40;n2=20;n3=40;
mu1=1;mu2=2;mu3=3;
sigma1=0.5;sigma2=0.5;sigma3=0.5;
iteration=300;*simulation times;
success=0;

*find real VUS using Gaussian quadrature;
LL1=mu1-3*sigma1;UL1=mu1+3*sigma1;
LL2=mu2-3*sigma2;UL2=mu2+3*sigma2;
LL3=mu3-3*sigma3;UL3=mu3+3*sigma3;
w1=sqrt(5-2*sqrt(10/7))/3;w2=-sqrt(5-2*sqrt(10/7))/3;
w3=sqrt(5+2*sqrt(10/7))/3;w4=-sqrt(5+2*sqrt(10/7))/3;w5=0;
k11=(UL1-LL1)/2*w1+(LL1+UL1)/2;k12=(UL1-LL1)/2*w2+(LL1+UL1)/2;
k13=(UL1-LL1)/2*w3+(LL1+UL1)/2;k14=(UL1-LL1)/2*w4+(LL1+UL1)/2;k15=(UL1-
LL1)/2*w5+(LL1+UL1)/2;
k21=(UL2-LL2)/2*w1+(LL2+UL2)/2;k22=(UL2-LL2)/2*w2+(LL2+UL2)/2;
k23=(UL2-LL2)/2*w3+(LL2+UL2)/2;k24=(UL2-LL2)/2*w4+(LL2+UL2)/2;k25=(UL2-
LL2)/2*w5+(LL2+UL2)/2;
k31=(UL3-LL3)/2*w1+(LL3+UL3)/2;k32=(UL3-LL3)/2*w2+(LL3+UL3)/2;
k33=(UL3-LL3)/2*w3+(LL3+UL3)/2;k34=(UL3-LL3)/2*w4+(LL3+UL3)/2;k35=(UL3-
LL3)/2*w5+(LL3+UL3)/2;
f1=a2**2/((a1+a2)*(a2+a3))*probnorm((k21-mu1)/sigma1)*(1-probnorm((k21-
mu3)/sigma3))*pdf('normal',k21,mu2,sigma2)*(UL2-LL2)/2
+a2*a3/(2*(a1+a2)*(a2+a3))*(1-probnorm((k11-mu2)/sigma2))*(1-probnorm((k11-
mu3)/sigma3))*pdf('normal',k11,mu1,sigma1)*(UL1-LL1)/2
+a1*a2/(2*(a1+a2)*(a2+a3))*probnorm((k31-mu1)/sigma1)*probnorm((k31-
mu2)/sigma2)*pdf('normal',k31,mu3,sigma3)*(UL3-LL3)/2;
f2=a2**2/((a1+a2)*(a2+a3))*probnorm((k22-mu1)/sigma1)*(1-probnorm((k22-
mu3)/sigma3))*pdf('normal',k22,mu2,sigma2)*(UL2-LL2)/2
```

```
+a2*a3/(2*(a1+a2)*(a2+a3))*(1-probnorm((k12-mu2)/sigma2))*(1-probnorm((k12-
mu3)/sigma3))*pdf('normal',k12,mu1,sigma1)*(UL1-LL1)/2
+a1*a2/(2*(a1+a2)*(a2+a3))*probnorm((k32-mu1)/sigma1)*probnorm((k32-
mu2)/sigma2)*pdf('normal',k32,mu3,sigma3)*(UL3-LL3)/2;
f3=a2**2/((a1+a2)*(a2+a3))*probnorm((k23-mu1)/sigma1)*(1-probnorm((k23-
mu3)/sigma3))*pdf('normal',k23,mu2,sigma2)*(UL2-LL2)/2
+a2*a3/(2*(a1+a2)*(a2+a3))*(1-probnorm((k13-mu2)/sigma2))*(1-probnorm((k13-
mu3)/sigma3))*pdf('normal',k13,mu1,sigma1)*(UL1-LL1)/2
+a1*a2/(2*(a1+a2)*(a2+a3))*probnorm((k33-mu1)/sigma1)*probnorm((k33-
mu2)/sigma2)*pdf('normal',k33,mu3,sigma3)*(UL3-LL3)/2;
f4=a2**2/((a1+a2)*(a2+a3))*probnorm((k24-mu1)/sigma1)*(1-probnorm((k24-
mu3)/sigma3))*pdf('normal',k24,mu2,sigma2)*(UL2-LL2)/2
+a2*a3/(2*(a1+a2)*(a2+a3))*(1-probnorm((k14-mu2)/sigma2))*(1-probnorm((k14-
mu3)/sigma3))*pdf('normal',k14,mu1,sigma1)*(UL1-LL1)/2
+a1*a2/(2*(a1+a2)*(a2+a3))*probnorm((k34-mu1)/sigma1)*probnorm((k34-
mu2)/sigma2)*pdf('normal',k34,mu3,sigma3)*(UL3-LL3)/2;
f5=a2**2/((a1+a2)*(a2+a3))*probnorm((k25-mu1)/sigma1)*(1-probnorm((k25-
mu3)/sigma3))*pdf('normal',k25,mu2,sigma2)*(UL2-LL2)/2
+a2*a3/(2*(a1+a2)*(a2+a3))*(1-probnorm((k15-mu2)/sigma2))*(1-probnorm((k15-
mu3)/sigma3))*pdf('normal',k15,mu1,sigma1)*(UL1-LL1)/2
+a1*a2/(2*(a1+a2)*(a2+a3))*probnorm((k35-mu1)/sigma1)*probnorm((k35-
mu2)/sigma2)*pdf('normal',k35,mu3,sigma3)*(UL3-LL3)/2;
VUS=(f1+f2)*(322+13*sqrt(70))/900+(f3+f4)*(322-
13*sqrt(70))/900+f5*128/225+a1*a3/(4*(a1+a2)*(a2+a3))*probnorm((mu3-
mu1)/sqrt(sigma1**2+sigma3**2));

************************************************;
*U-statistics;
************************************************;

do p=1 to iteration;

*get random number;
x=j(n1,1,0);
y=j(n2,1,0);
z=j(n3,1,0);
do i=1 to n1;
x[i,]=rand('normal',mu1,sigma1);
end;
do i=1 to n2;
y[i,]=rand('normal',mu2,sigma2);
end;
do i=1 to n3;
z[i,]=rand('normal',mu3,sigma3);
end;

*find estimate VUS and variance using non-parametric method;
VUS_hat=0;sigma111=0;sigma100=0;sigma010=0;sigma001=0;sigma110=0;sigma101=0;si
gma011=0;
do i=1 to n1;
do j=1 to n2;
do k=1 to n3;
if x[i,]<=y[j,] then s1=1;else s1=0;
if y[j,]<=z[k,] then s2=1;else s2=0;
if x[i,]<=z[k,] then s3=1;else s3=0;

do l=1 to n2;
do m=1 to n3;
if x[i,]<=y[l,] then s_11=1;else s_11=0;
```

```
if y[l,]<=z[m,] then s_12=1;else s_12=0;
if x[i,]<=z[m,] then s_13=1;else s_13=0;
sigma100=sigma100+(a2/(a2+a3)*s2+a3/(2*(a2+a3)))*(a2/(a1+a2)*s1+a1/(2*(a1+a2))
)*s3*(a2/(a2+a3)*s_12+a3/(2*(a2+a3)))*(a2/(a1+a2)*s_11+a1/(2*(a1+a2)))*s_13;
end;
end;

do l=1 to n1;
do m=1 to n3;
if x[l,]<=y[j,] then s_21=1;else s_21=0;
if y[j,]<=z[m,] then s_22=1;else s_22=0;
if x[l,]<=z[m,] then s_23=1;else s_23=0;
sigma010=sigma010+(a2/(a2+a3)*s2+a3/(2*(a2+a3)))*(a2/(a1+a2)*s1+a1/(2*(a1+a2))
)*s3*(a2/(a2+a3)*s_22+a3/(2*(a2+a3)))*(a2/(a1+a2)*s_21+a1/(2*(a1+a2)))*s_23;
end;
end;

do l=1 to n1;
do m=1 to n2;
if x[l,]<=y[m,] then s_31=1;else s_31=0;
if y[m,]<=z[k,] then s_32=1;else s_32=0;
if x[l,]<=z[k,] then s_33=1;else s_33=0;
sigma001=sigma001+(a2/(a2+a3)*s2+a3/(2*(a2+a3)))*(a2/(a1+a2)*s1+a1/(2*(a1+a2))
)*s3*(a2/(a2+a3)*s_32+a3/(2*(a2+a3)))*(a2/(a1+a2)*s_31+a1/(2*(a1+a2)))*s_33;
end;
end;

do l=1 to n3;
if x[i,]<=y[j,] then s_41=1;else s_41=0;
if y[j,]<=z[l,] then s_42=1;else s_42=0;
if x[i,]<=z[l,] then s_43=1;else s_43=0;
sigma110=sigma110+(a2/(a2+a3)*s2+a3/(2*(a2+a3)))*(a2/(a1+a2)*s1+a1/(2*(a1+a2))
)*s3*(a2/(a2+a3)*s_42+a3/(2*(a2+a3)))*(a2/(a1+a2)*s_41+a1/(2*(a1+a2)))*s_43;
end;

do l=1 to n2;
if x[i,]<=y[l,] then s_51=1;else s_51=0;
if y[l,]<=z[k,] then s_52=1;else s_52=0;
if x[i,]<=z[k,] then s_53=1;else s_53=0;
sigma101=sigma101+(a2/(a2+a3)*s2+a3/(2*(a2+a3)))*(a2/(a1+a2)*s1+a1/(2*(a1+a2))
)*s3*(a2/(a2+a3)*s_52+a3/(2*(a2+a3)))*(a2/(a1+a2)*s_51+a1/(2*(a1+a2)))*s_53;
end;

do l=1 to n1;
if x[l,]<=y[j,] then s_61=1;else s_61=0;
if y[j,]<=z[k,] then s_62=1;else s_62=0;
if x[l,]<=z[k,] then s_63=1;else s_63=0;
sigma011=sigma011+(a2/(a2+a3)*s2+a3/(2*(a2+a3)))*(a2/(a1+a2)*s1+a1/(2*(a1+a2))
)*s3*(a2/(a2+a3)*s_62+a3/(2*(a2+a3)))*(a2/(a1+a2)*s_61+a1/(2*(a1+a2)))*s_63;
end;

VUS_hat=VUS_hat+(a2/(a2+a3)*s2+a3/(2*(a2+a3)))*(a2/(a1+a2)*s1+a1/(2*(a1+a2)))*
s3;
sigma111=sigma111+((a2/(a2+a3)*s2+a3/(2*(a2+a3)))*(a2/(a1+a2)*s1+a1/(2*(a1+a2)
))*s3)**2;

end;
end;
end;
```

```
VUS_hat=VUS_hat/(n1*n2*n3);
sigma111=sigma111/(n1*n2*n3);
sigma100=sigma100/(n1*n2*n3*n2*n3);
sigma010=sigma010/(n1*n2*n3*n1*n3);
sigma001=sigma001/(n1*n2*n3*n1*n2);
sigma110=sigma110/(n1*n2*n3*n3);
sigma101=sigma101/(n1*n2*n3*n2);
sigma011=sigma011/(n1*n2*n3*n1);
variance=(sigma111-VUS_hat**2)/(n1*n2*n3)+(sigma100-VUS_hat**2)*(n2-1)*(n3-
1)/(n1*n2*n3)
       +(sigma010-VUS_hat**2)*(n1-1)*(n3-1)/(n1*n2*n3)+(sigma001-
VUS_hat**2)*(n1-1)*(n2-1)/(n1*n2*n3)
       +(sigma110-VUS_hat**2)*(n3-1)/(n1*n2*n3)+(sigma101-VUS_hat**2)*(n2-
1)/(n1*n2*n3)
       +(sigma011-VUS_hat**2)*(n1-1)/(n1*n2*n3);

*find the coverage rate;
Low=VUS_hat-1.96*sqrt(variance);
High=VUS_hat+1.96*sqrt(variance);
if VUS<=High & VUS>=Low then success=success+1;
end;
coverage_rate=success/iteration;
print success coverage_rate VUS;
quit;

************************************************;
*bootstrap;
************************************************;
do p=1 to iteration;

*get random number;
x=j(n1,1,0);
y=j(n2,1,0);
z=j(n3,1,0);
do i=1 to n1;
x[i,]=rand('normal',mu1,sigma1);
end;
do i=1 to n2;
y[i,]=rand('normal',mu2,sigma2);
end;
do i=1 to n3;
z[i,]=rand('normal',mu3,sigma3);
end;

*find estimate VUS and Confidence interval using Bootstrap method;
VUS_hat=0;
do i=1 to n1;
do j=1 to n2;
do k=1 to n3;
if x[i,]<=y[j,] then s1=1;else s1=0;
if y[j,]<=z[k,] then s2=1;else s2=0;
if x[i,]<=z[k,] then s3=1;else s3=0;
VUS_hat=VUS_hat+(a2/(a2+a3)*s2+a3/(2*(a2+a3)))*(a2/(a1+a2)*s1+a1/(2*(a1+a2)))*
s3;
end;
end;
end;
VUS_hat=VUS_hat/(n1*n2*n3);
```

```
VUS_resample=j(200,1,0);
do b=1 to 200;
x_b=j(n1,1,0);
y_b=j(n2,1,0);
z_b=j(n3,1,0);
do i=1 to n1;
r_number=int(rand('Uniform')*n1)+1;
x_b[i,]=x[r_number,];
end;
do j=1 to n2;
r_number=int(rand('Uniform')*n2)+1;
y_b[j,]=y[r_number,];
end;
do k=1 to n3;
r_number=int(rand('Uniform')*n3)+1;
z_b[k,]=z[r_number,];
end;
do i=1 to n1;
do j=1 to n2;
do k=1 to n3;
if x_b[i,]<=y_b[j,] then s1=1;else s1=0;
if y_b[j,]<=z_b[k,] then s2=1;else s2=0;
if x_b[i,]<=z_b[k,] then s3=1;else s3=0;
VUS_resample[b,]=VUS_resample[b,]+(a2/(a2+a3)*s2+a3/(2*(a2+a3)))*(a2/(a1+a2)*s
1+a1/(2*(a1+a2)))*s3;
end;
end;
end;
VUS_resample[b,]=VUS_resample[b,]/(n1*n2*n3);
end;
VUS_resample1=VUS_resample;
VUS_resample[rank(VUS_resample),]=VUS_resample1;

*find the coverage rate;
Low=VUS_resample[5,];
High=VUS_resample[195,];
if VUS<=High & VUS>=Low then success=success+1;
end;
coverage_rate=success/iteration;
print success coverage_rate VUS;
quit;
```

descriptive Study for Composite IQ. For other variables, the code is similar

```
proc import datafile='D:\My Documents\Mixed model\SB_long_yusheng.xls'
out=alldata replace;
run;
proc import datafile='D:\My Documents\Mixed model\SB_long_yusheng_adj.xls'
out=adjdata replace;
run;
data alldata;
set alldata;
keep idnum age_diagnosis_years time_since_diagnosis_years sex SES
Radiation Chemotherapy NPS_Total vrsas avrsas qrsas stmsas compiq;
rename age_diagnosis_years=age time_since_diagnosis_years=time;
run;
```

```
data adjdata;
set adjdata;
keep idnum age_diagnosis_years time_since_diagnosis_years sex SES
Radiation Chemotherapy NPS_Total vrsas avrsas qrsas stmsas compiq;
rename age_diagnosis_years=age time_since_diagnosis_years=time;
run;
data alldata;
set alldata;
age_group=age>7;
run;
data adjdata;
set adjdata;
age_group=age>7;
run;
ods output Freq.Table1.OneWayFreqs=d1;
proc freq data=alldata;
table idnum;
run;
ods output Freq.Table1.OneWayFreqs=d2;
proc freq data=adjdata;
table idnum;
run;
proc means data=adjdata noprint;
var radiation chemotherapy sex age_group SES;
by idnum;
output out=mean;
run;
data mean;
set mean;
if _STAT_='MEAN';
radiation=(radiation>0);
chemotherapy=(chemotherapy>0);
run;
proc freq data=adjdata;
table SES;
run;
proc univariate data=adjdata;
var NPS_total age time;
run;
data d2;
set d2;
if frequency=2 then f=2;
else if frequency=3 then f=3;
else if frequency=4 then f=4;
else f=5;
run;
proc freq data=d2;
table f;
run;

*****************************************
Individual Composite IQ score trajectories
*****************************************;

goption reset=all;
symbol interpol=join  repeat=300  ;
proc gplot data=adjdata;
    plot compiq* time=idnum;
    run;quit;
```

Model construction, WAPRESS and VUS for Composite IQ NPS hierarchical linear model. For other linear/quatratic models, the code is similar

```
*****************************
Model construction and WAPRESS
*****************************;

proc import datafile='D:\My documents\Mixed model\SB_long_yusheng_adj.xls'
out=alldata replace;
run;
data alldata;
set alldata;
keep idnum age_diagnosis_years time_since_diagnosis_years sex SES
Radiation Chemotherapy NPS_Total vrsas avrsas qrsas stmsas compiq;
rename age_diagnosis_years=age time_since_diagnosis_years=time NPS_total=NPS
Chemotherapy=chemo Radiation=rad;
run;
data alldata;
set alldata;
age_g=age>7;
time2=time**2;
compiq_=compiq;
run;

data alldata;
retain index 0;
set alldata;by idnum;
if first.idnum then index=index+1;
run;

data out;
set _null_;
run;

%macro press;
%do i=1 %to 95;

data data&i;
set alldata;
if index=&i then compiq=.;
run;
ods output  SolutionF=est&i;
proc mixed data=data&i covtest;
         model compiq= age_g SES NPS time
                    NPS*time
                /s outpred=out&i;
         random intercept time/s  type=un;
         run;
data out&i;
set out&i;
if index=&i;
run;
data out;
set out out&i;
run;
%end;
```

```
%mend;
%press;
data out;
set out;;
se=(compiq_-pred)**2;
run;
proc means data=out noprint;
var se;
by idnum;
output out=out_1 mean=average;
run;
proc means sum data=out_1;
var average;
run;

*****************************
VUS calculation
*****************************;

proc import datafile='D:\My Documents\Mixed model\SB_long_yusheng_adj.xls'
out=alldata replace;
run;
data alldata;
set alldata;
keep idnum age_diagnosis_years time_since_diagnosis_years sex SES
Radiation Chemotherapy NPS_Total vrsas avrsas qrsas stmsas compiq;
rename age_diagnosis_years=age time_since_diagnosis_years=time NPS_total=NPS
Chemotherapy=chemo Radiation=rad;
run;
data alldata;
set alldata;
age_g=age>7;
time2=time**2;
run;
ods output SolutionF=est;
proc mixed data=alldata covtest;
        model compiq= ses age_g NPS time
                    NPS*time
                /s outpred=out1;
        random intercept time/s  type=un subject=idnum;
        run;

data est;
set est (keep=effect estimate);
run;
proc transpose data=est out=estimate;
var estimate;
id effect;
run;
data out;
set alldata;
_name_="Estimate";
run;
data estimate;
set estimate;
rename ses=ses_e age_g=age_g_e NPS=NPS_e time=time_e;
run;
data out;
merge out estimate;
by _name_;
```

```
run;
data out;
set out;
y=intercept+ses_e*ses+age_g_e*age_g+NPS_e*NPS+NPS_time*NPS*time+time_e*time;
run;

***************
3-level VUS
***************;
proc iml;
use out var{compiq y};
read all var{y} where(compiq<86) into x;
read all var{y} where(compiq>=86 & compiq<101) into y;
read all var{y} where(compiq>=101) into z;
close out;
n1=nrow(x);n2=nrow(y);n3=nrow(z);
s=0;
do i=1 to n1;
do j=1 to n2;
do k=1 to n3;
a=0;b=0;c=0;
if x[i,]<=y[j,] then a=1;
if y[j,]<=z[k,] then b=1;
if x[i,]<=z[k,] then c=1;
s=s+c*(a/2+1/4)*(b/2+1/4);
end;
end;
end;
vus=s/(n1*n2*n3);
print vus;
create vus var{vus};
append;
close vus;
quit;

***************
4-level VUS
***************;
proc iml;
use out var{compiq y};
read all var{y} where(compiq<82) into x1;
read all var{y} where(compiq>=82 & compiq<94) into x2;
read all var{y} where(compiq>=94 & compiq<105) into x3;
read all var{y} where(compiq>=105) into x4;
close out;
n1=nrow(x1);n2=nrow(x2);n3=nrow(x3);n4=nrow(x4);
sum=0;
do i=1 to n1;
do j=1 to n2;
do k=1 to n3;
do l=1 to n4;
a=(x1[i,]<=x2[j,]);
b=(x1[i,]<=x3[k,]);
c=(x1[i,]<=x4[l,]);
d=(x2[j,]<=x3[k,]);
e=(x2[j,]<=x4[l,]);
f=(x3[k,]<=x4[l,]);
sum=sum+c*(3*a*d*f+a*f+(b*f+a*e+a*d*e+b*d*f)/2+(a+f+a*d+d*f)/3+b*e/4+(b+e)/6+1
/9)/36;
end;
```

```
end;
end;
end;
vus=sum/(n1*n2*n3*n4);
print vus;
create vus var{vus};
append;
close vus;
quit;
```

Model construction, WAPRESS and VUS for Composite IQ NPS Compertz model. For other linear/quatratic models, the code is similar

```
*****************************
Model construction and WAPRESS
*****************************;

proc import datafile='H:\Mixed model\SB_long_yusheng_adj.xls' out=alldata
replace;
run;
data alldata;
set alldata;
keep idnum age_diagnosis_years time_since_diagnosis_years sex SES
Radiation Chemotherapy NPS_Total vrsas avrsas qrsas stmsas compiq;
rename age_diagnosis_years=age time_since_diagnosis_years=time;
run;
data alldata;
set alldata;
age_group=age>7;
run;

data alldata;
retain index 0;
set alldata;by idnum;
if first.idnum then index=index+1;
run;

data estimate;
set _null_;
run;

%macro press;
%do i=1 %to 95;

data data&i;
set alldata;
if index=&i then delete;
run;
ods output  ParameterEstimates=est&i;
proc nlmixed data=data&i;
parms beta01=0.5 beta11=0.5 beta61=0.5
     beta32=0.5 beta62=0.5
```

```
     beta13=0.5 beta33=0.5
     sigmae=.5 sigmau1=.5 sigmau2=.5 sigmau12=0;
b1=beta01+age_group*beta11+NPS_total*beta61+u1;
b2=SES*beta32+NPS_total*beta62+u2;
b3=age_group*beta13+SES*beta33;
yhat=b1-b2*exp(-b3*time);
compiqhat=exp(yhat);
y=log(compiq);
model y~normal(yhat,sigmae);
random u1 u2~normal([0,0],[sigmau1,sigmau12,sigmau2]) subject=idnum;
run;
data est&i;
set est&i (keep=parameter estimate);
run;
proc transpose data=est&i out=estimate&i;
var estimate;
id parameter;
run;
data estimate&i;
set estimate&i;
index=&i;
run;
data estimate;
set estimate estimate&i;
run;
%end;
%mend;
%press;
data alldata;
merge alldata estimate;
by index;
run;

data alldata;
set alldata;
b1=beta01+age_group*beta11+NPS_total*beta61;
b2=SES*beta32+NPS_total*beta62;
b3=age_group*beta13+SES*beta33;
y=exp(b1-b2*exp(-b3*time)+0.5*(sigmau1+sigmau2*exp(-2*b3*time)-2*exp(-
b3*time)*sigmau12+sigmae));
se=(y-compiq)**2;
run;
proc means data=alldata noprint;
var se;
by idnum;
output out=out_1 mean=average;
run;
proc means sum data=out_1;
var average;
run;
*****************************
VUS calculation
*****************************;

proc import datafile='D:\My documents\Mixed model\SB_long_yusheng_adj.xls'
out=alldata replace;
run;
data alldata;
set alldata;
keep idnum age_diagnosis_years time_since_diagnosis_years sex SES
```

```
Radiation Chemotherapy NPS_Total vrsas avrsas qrsas stmsas compiq;
rename age_diagnosis_years=age time_since_diagnosis_years=time NPS_total=NPS
Chemotherapy=chemo Radiation=rad;
run;
data alldata;
set alldata;
age_g=age>7;
run;
ods output  ParameterEstimates=est;
proc nlmixed data=alldata method=firo;
parms beta01=0.5 beta11=0.5 beta21=0.5          beta41=0.5
     beta02=0.5          beta22=0.5          beta42=0.5
                        beta23=0.5 beta33=0.5 beta43=0.5
     sigmae=.5 sigmau1=.5;
b1=beta01+age_g*beta11+SES*beta21          +NPS*beta41+u1;
b2=beta02           +SES*beta22          +NPS*beta42;
b3=                 SES*beta23+sex*beta33+NPS*beta43;
yhat=b1-b2*exp(-b3*time);
compiqhat=exp(yhat);
y=log(compiq);
model y~normal(yhat,sigmae);
random u1~normal(0,sigmau1) subject=idnum;
predict compiqhat out=out;
run;
data est;
set est (keep=parameter estimate);
run;
proc transpose data=est out=estimate;
var estimate;
id parameter;
run;
data out;
set out;
_name_="Estimate";
run;
data out;
merge out estimate;
by _name_;
run;
data out;
set out;
b0=beta01+age_g*beta11+SES*beta21          +NPS*beta41;
b1=beta02           +SES*beta22          +NPS*beta42;
b2=                 SES*beta23+sex*beta33+NPS*beta43;
y=exp(b0-b1*exp(-b2*time)+0.5*(sigmau1+sigmae));
run;
***************
3-level VUS
***************;
proc iml;
use out var{compiq y};
read all var{y} where(compiq<86) into x;
read all var{y} where(compiq>=86 & compiq<101) into y;
read all var{y} where(compiq>=101) into z;
close out;
n1=nrow(x);n2=nrow(y);n3=nrow(z);
s=0;
do i=1 to n1;
do j=1 to n2;
do k=1 to n3;
```

```
a=0;b=0;c=0;
if x[i,]<=y[j,] then a=1;
if y[j,]<=z[k,] then b=1;
if x[i,]<=z[k,] then c=1;
s=s+c*(a/2+1/4)*(b/2+1/4);
end;
end;
end;
vus=s/(n1*n2*n3);
print vus;
create vus var{vus};
append;
close vus;
quit;

***************
4-level VUS
***************;
proc iml;
use out var{compiq y};
read all var{y} where(compiq<82) into x1;
read all var{y} where(compiq>=82 & compiq<94) into x2;
read all var{y} where(compiq>=94 & compiq<105) into x3;
read all var{y} where(compiq>=105) into x4;
close out;
n1=nrow(x1);n2=nrow(x2);n3=nrow(x3);n4=nrow(x4);
sum=0;
do i=1 to n1;
do j=1 to n2;
do k=1 to n3;
do l=1 to n4;
a=(x1[i,]<=x2[j,]);
b=(x1[i,]<=x3[k,]);
c=(x1[i,]<=x4[l,]);
d=(x2[j,]<=x3[k,]);
e=(x2[j,]<=x4[l,]);
f=(x3[k,]<=x4[l,]);
sum=sum+c*(3*a*d*f+a*f+(b*f+a*e+a*d*e+b*d*f)/2+(a+f+a*d+d*f)/3+b*e/4+(b+e)/6+1
/9)/36;
end;
end;
end;
end;
vus=sum/(n1*n2*n3*n4);
print vus;
create vus var{vus};
append;
close vus;
quit;
```

```
#####################################################################
R-code for one example of Empircal likelihood Inference for VUS
#####################################################################
#simulation parameter

Sim=1                    #simulation number
loop=10                  #solving equation system loop number
VUS=0.3029789            #Theoritical VUS
a1=2
a2=1
a3=2
n1=16       #sample size X
n2=8        #sample size Y
n3=16       #sample size Z


# construct necessary functions

H=function(x1,x2,x3,theta)
(a2^2*(x1<=x2)*(x2<=x3)+a2*a3/2*(x1<=x2)*(x1<=x3)+a1*a2/2*(x1<=x3)*(x2<=x3)+a1*a3/4*(x1<=x3))/((a2+a
3)*(a1+a2))-theta

Hi..=function(u2,u3,theta){
S=rep(0,n1)
for(i in 1:n1)for(q in 1:n2)for(r in 1:n3)S[i]=S[i]+u2[q]*u3[r]*H(x[i],y[q],z[r],theta)
return(S)
}
H.j.=function(u1,u3,theta){
S=rep(0,n2)
for(j in 1:n2)for(p in 1:n1)for(r in 1:n3)S[j]=S[j]+u1[p]*u3[r]*H(x[p],y[j],z[r],theta)
return(S)
}
H..k=function(u1,u2,theta){
S=rep(0,n3)
for(k in 1:n3)for(p in 1:n1)for(q in 1:n2)S[k]=S[k]+u1[p]*u2[q]*H(x[p],y[q],z[k],theta)
return(S)
}
lamda.restrict=function(u1,u2,u3,theta){
S=0
for(i in 1:n1)for(j in 1:n2)for(k in 1:n3)S=S+u1[i]*u2[j]*u3[k]*H(x[i],y[j],z[k],theta)
return(S)
}
log.lik=function(u1,u2,u3,lamda,theta)
2*sum(log(abs(1+lamda/n1*Hi..(u2,u3,theta))))+2*sum(log(abs(1+lamda/n2*H.j.(u1,u3,theta))))+2*sum(log(abs(1+
lamda/n3*H..k(u1,u2,theta))))

g1=function(w)
sum((w[1:n1]-
1/(n1+w[n1+n2+n3+1]*Hi..(w[(n1+1):(n1+n2)],w[(n1+n2+1):(n1+n2+n3)],w[n1+n2+n3+2])))^2)+sum((w[(n1+1):(
n1+n2)]-
1/(n2+w[n1+n2+n3+1]*H.j.(w[1:n1],w[(n1+n2+1):(n1+n2+n3)],w[n1+n2+n3+2])))^2)+sum((w[(n1+n2+1):(n1+n2+
n3)]-
1/(n3+w[n1+n2+n3+1]*H..k(w[1:n1],w[(n1+1):(n1+n2)],w[n1+n2+n3+2])))^2)+lamda.restrict(w[1:n1],w[(n1+1):(n
```

```
1+n2)],w[(n1+n2+1):(n1+n2+n3)],w[n1+n2+n3+2])^2+(log.lik(w[1:n1],w[(n1+1):(n1+n2)],w[(n1+n2+1):(n1+n2+n
3)],w[n1+n2+n3+1],w[n1+n2+n3+2])-qchisq(0.95,1))^2


Cover=0
length=0

for(q in 1:Sim){
x=rnorm(n1,mean=1,sd=1.25)
y=rnorm(n2,mean=2,sd=1.25)
z=rnorm(n3,mean=3,sd=1.25)

#find estimate VUS
VUS_hat=0
for( i in 1:n1)
for( j in 1:n2)
for( k in 1:n3)
{
s1=x[i]<=y[j]
s2=y[j]<=z[k]
s3=x[i]<=z[k]
VUS_hat=VUS_hat+(a2/(a2+a3)*s2+a3/(2*(a2+a3)))*(a2/(a1+a2)*s1+a1/(2*(a1+a2)))*s3
}
VUS_hat=VUS_hat/(n1*n2*n3)

# estimate confidence limit

NA.LL=VUS_hat-1.96*0.03
NA.UL=VUS_hat+1.96*0.03

##################Empirical likelihood#################

#construct EL CI by solving nonlinear equation system


u1=rep(1/n1,n1)
u2=rep(1/n2,n2)
u3=rep(1/n3,n3)
EL.LL=rep(NA,loop)
parameter=c(0,NA.LL)
for (p in 1:loop){
H1=Hi..(u2,u3,0)
H11=0
for(j in 1:n2)for(k in 1:n3)H11=H11+u2[j]*u3[k]
H2=H.j.(u1,u3,0)
H22=0
for(i in 1:n1)for(k in 1:n3)H22=H22+u1[i]*u3[k]
H3=H..k(u1,u2,0)
H33=0
for(i in 1:n1)for(j in 1:n2)H33=H33+u1[i]*u2[j]
H4=lamda.restrict(u1,u2,u3,0)
H44=0
for(i in 1:n1)for(j in 1:n2)for(k in 1:n3)H44=H44+u1[i]*u2[j]*u3[k]
g2=function(par)(2*sum(log(abs(1+par[1]/n1*(H1-par[2]*H11))))+2*sum(log(abs(1+par[1]/n2*(H2-
par[2]*H22))))+2*sum(log(abs(1+par[1]/n3*(H3-par[2]*H33)))))-qchisq(0.95,1))^2+(H4-par[2]*H44)^2
```

```
parameter=nlm(g2,parameter)$estimate
EL.LL[p]=parameter[2]
u1=1/(n1+parameter[1]*Hi..(u2,u3,parameter[2]))
u2=1/(n2+parameter[1]*H.j.(u1,u3,parameter[2]))
u3=1/(n3+parameter[1]*H..k(u1,u2,parameter[2]))
}
ELT.LL=EL.LL[loop] #EL lower limit

u1=rep(1/n1,n1)
u2=rep(1/n2,n2)
u3=rep(1/n3,n3)
EL.UL=rep(NA,loop)
parameter=c(0,NA.UL)
for (p in 1:loop){
H1=Hi..(u2,u3,0)
H11=0
for(j in 1:n2)for(k in 1:n3)H11=H11+u2[j]*u3[k]
H2=H.j.(u1,u3,0)
H22=0
for(i in 1:n1)for(k in 1:n3)H22=H22+u1[i]*u3[k]
H3=H..k(u1,u2,0)
H33=0
for(i in 1:n1)for(j in 1:n2)H33=H33+u1[i]*u2[j]
H4=lamda.restrict(u1,u2,u3,0)
H44=0
for(i in 1:n1)for(j in 1:n2)for(k in 1:n3)H44=H44+u1[i]*u2[j]*u3[k]
g2=function(par)(2*sum(log(abs(1+par[1]/n1*(H1-par[2]*H11))))+2*sum(log(abs(1+par[1]/n2*(H2-
par[2]*H22))))+2*sum(log(abs(1+par[1]/n3*(H3-par[2]*H33)))))-qchisq(0.95,1))^2+(H4-par[2]*H44)^2
parameter=nlm(g2,parameter)$estimate
EL.UL[p]=parameter[2]
u1=1/(n1+parameter[1]*Hi..(u2,u3,parameter[2]))
u2=1/(n2+parameter[1]*H.j.(u1,u3,parameter[2]))
u3=1/(n3+parameter[1]*H..k(u1,u2,parameter[2]))
}
ELT.UL=EL.UL[loop] #EL upper limit

length=length+(ELT.UL-ELT.LL)
if(ELT.LL<=VUS & VUS<=ELT.UL)Cover=Cover+1
}
Coverage=Cover/Sim
length=length/Sim
Coverage
length
```